

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 875 575 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:

**04.11.1998 Bulletin 1998/45**

(51) Int Cl.<sup>6</sup>: **C12N 15/54, C12N 9/10**

(21) Application number: **98302489.4**

(22) Date of filing: **31.03.1998**

(84) Designated Contracting States:

**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**

Designated Extension States:

**AL LT LV MK RO SI**

• Ihara, Yuri

**Takaishi, Osaka (JP)**

• Hayashi, Takahisa

**Uji Kyoto 611 (JP)**

(30) Priority: **01.04.1997 JP 83133/97**

(71) Applicant: **NISSHINBO INDUSTRIES, INC.**

**Chuo-ku, Tokyo 103 (JP)**

(74) Representative: **Bannerman, David Gardner et al**

**Withers & Rogers**

**4 Dyer's Buildings**

**Holborn**

**London, EC1N 2JT (GB)**

(72) Inventors:

• **Hasegawa, Osamu, c/o Nisshinbo Ind., Inc.  
Adachi-ku, Tokyo (JP)**

• **Aotsuka, Satoshi c/o Nisshinbo Ind., Inc.,  
Adachi-ku, Tokyo (JP)**

Remarks:

The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) **Cellulose synthase gene**

(57) mRNA was extracted at the stage for cotton plant fibrous cells to accumulate cellulose, and cDNA's complementary thereto were synthesized to construct a cDNA library. Clones of a number of 750 were arbitrarily selected from the library, and they were randomly subjected from to sequencing. Those having homology to

an amino acid sequence deduced from a gene of cellulose 4- $\beta$ -glucosyltransferase (bcsA) of cellulose synthase operon of acetic acid bacterium were selected from obtained nucleotide sequences of the respective clones. Thus, DNA coding for cellulose synthase was obtained.

**EP 0 875 575 A2**

**Description**Technical Field

5 The present invention relates to a DNA coding for cellulose synthase originating from cotton plant (Gossypium hirsutum), a recombinant DNA containing the same, a transformed cell transformed with the DNA, and a method for controlling cellular cellulose synthesis.

Background Art

10 Cellulose is used for paper, woody structural materials, fiber, cloths, food, cosmetics, and pharmaceuticals, as well as it is utilized as energy. Therefore, cellulose is industrially useful and valuable. Cellulose is capable of forming a variety of crystalline structures, and hence it is expected to develop a new material by controlling enzymes involved in biosynthesis of cellulose. The cellulose-related industry has been hitherto directed to such cellulose products that  
 15 have been already produced, in which there has been no trial to develop a new material based on an aspect of biosynthesis. The mechanism of disease action, which is exerted by pathogenic microorganisms on plants, often results from the inhibition on cellulose biosynthesis as in Pyricularia oryzae (P. oryzae). Therefore, the addition of disease resistance to the cellulose biosynthesis mechanism is agriculturally applicable and valuable. Further, cellulose is the most abundant organic compound on the earth, and it is a sink in which the largest amount of CO<sub>2</sub> in the atmospheric  
 20 air is fixed. Therefore, the genetic improvement of cellulose biosynthesis enzymes is also applicable to the industry which is directed to the control of CO<sub>2</sub> in the atmospheric air based on the use of cellulose as the sink.

In recent years, cDNA's originating from fiber cells of cotton plant have been randomly sequenced, and it has been reported that full length CelA1 and partial length of CelA2 probably represent cDNAs of cotton plant cellulose synthase, in view of the homology to bacterial cellulose synthase gene (bacterial BcsA) (Pear et al., Proceeding of National  
 25 Academy of Science, USA (1996) 93 12637-12642). The binding ability to UDP-glucose has been demonstrated for CelA1. However, as for CelA2, the homology has been merely demonstrated for the C-terminal amino acid sequence.

Disclosure of the Invention

30 The present invention has been made in order to provide a new method for regulating cellulose production in prokaryotic cells or eukaryotic cells, an object of which is to provide a DNA coding for cellulose synthase, a recombinant DNA containing the same, a transformed cell transformed with the DNA, and a method for regulating cellular cellulose synthesis.

The present inventors firstly extracted mRNAs at the stage for cotton plant fiber cells to accumulate cellulose, and  
 35 cDNAs complementary thereto were synthesized to construct a cDNA library. 750 of cDNA clones were arbitrarily selected from the library, and they were randomly subjected to sequencing. Six amino acid sequences were derived for one nucleotide sequence of each of the obtained clones to select those having homology to an amino acid sequence obtained by translation from a gene of cellulose 4- $\beta$ -glucosyltransferase (bcsA) of cellulose synthase operon of aceto-  
 40 bacterium. As a result, genes, which were classified into three types or groups, were found, and they were designated as PcsA1, PcsA2, and PcsA3 respectively (PcsA is an abbreviation of "Plant Cellulose Synthase A").

That is, the present invention lies in a DNA coding for any one of the following proteins (A) to (C):

(A) a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO: 2 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids  
 45 relevant to SEQ ID NO: 2;

(B) a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO: 4 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO: 4; and

(C) a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO: 8 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO: 8, and comprising an amino acid sequence shown in SEQ ID NO: 11 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID  
 50 NO: 11.

55 In another aspect, the present invention provides a recombinant vector comprising all or a part of the DNA as defined above, and a transformed cell transformed with the DNA as defined above.

In still another aspect, the present invention provides a method for regulating cellulose synthesis in a cell, comprising the steps of introducing the DNA as defined above into the cell, and expressing RNA having a nucleotide

sequence homologous to the DNA as defined above or a nucleotide sequence complementary to the DNA as defined above.

SEQ ID NO: 1 corresponds to a sequence of PcsA1, and SEQ ID NO: 3 corresponds to a sequence of PcsA2. SEQ ID NO: 5 corresponds to a sequence of 3'-side region of PcsA3, SEQ ID NO: 7 corresponds to a sequence of 5'-side region of PcsA3, and SEQ ID NO: 9 corresponds to a sequence of internal region of PcsA3.

It has been demonstrated that PcsA1 and PcsA2 of the DNA's described above are DNA's coding for cotton plant cellulose synthase, according to the expression in eukaryotic cells (animal cells and/or yeast). It has been also demonstrated that an antibody thereagainst inhibits the cotton plant cellulose synthase activity in a cell-free system. Further, PcsA3, which is different from PcsA1 and PcsA2, has been found. Any one of these species was obtained as partial one, at the stage of clones obtained by the random sequencing, and no 5'-portion of the coding region was contained. Therefore, clones which have sequences of 5'-portions were isolated in accordance with the 5'-RACE method based on the use of PCR to determine the sequences. As a result of this operation, the sequences of the 5'-portions corresponding to the partial length clones were obtained for PcsA1 and PcsA2.

On the other hand, as for PcsA3, a sequence of a 5'-portion of another clone, which was considered to belong to the same PcsA3 group, was obtained. The both sequences had extremely high homology, and hence they were considered to have underwent multiple gene formation relatively recently originating from an identical gene through the process of duplication. Therefore, even when the both are combined with each other at corresponding portions to construct a fused gene followed by expression, it is assumed that the activity and function of a produced enzyme may not be affected thereby.

As for PcsA1 and PcsA2, in order to obtain a full length clone, primers were designed on the basis of the sequence of the 5'-portion and the sequence of the 3'-portion of the partial length clone to perform PCR. Thus, a clone containing ORF was obtained.

Those applicable as the template to be used for the RACE method may be any of cDNA synthesized from mRNA and a phage library. When the phage library is used, it is possible to use a sequence in the vector as a 5'-side primer.

As a result of random sequencing, seven clones concerning PcsA2 were most abundantly present, of 15 clones seemed to code the cellulose synthase. Expression was confirmed in eukaryotic cells (animal cells and/or yeast) transformed with the cellulose synthase gene. As a result, the cellulose synthase activity was observed.

The present invention will be explained in detail below.

#### <1> Preparation of cotton plant cDNA library

Cotton plant fiber cells at the stage of cellulose accumulation are preferably used as a material for extracting mRNA to construct a cotton plant cDNA library. The method for extracting mRNA is not specifically limited, for which it is possible to adopt an ordinary method for extracting mRNA from plant.

cDNA can be synthesized, for example, by using a poly T sequence which is complementary to poly A nucleotide existing at the terminal of mRNA as a primer to synthesize complementary DNA by the aid of reverse transcriptase, and forming a double strand by the aid of DNA polymerase.

The method therefor is described, for example, in Molecular Cloning (Maniatis et al., Cold Spring Harbour Laboratory). However, a variety of cDNA synthesis kits are commercially available from various companies, which may be used.

Generally, the library is constructed by using a phage vector. A variety of commercially available vectors are usable. However, it is preferable to use a vector, for example,  $\lambda$ ZAP vector in which it is unnecessary to perform recloning from the vector, and it is possible to immediately prepare a plasmid for sequencing.

#### <2> Determination of nucleotide sequence of cDNA

Clones are randomly selected from the obtained cDNA library to determine nucleotide sequences of inserts in the clones. The nucleotide sequence can be determined in accordance with the Maxam-Gilbert method or the dideoxy method. Among them, the dideoxy method is more convenient and preferred.

The nucleotide sequence can be determined in accordance with the dideoxy method by using a commercially available sequencing kit. Further, the use of an automatic sequencer makes it possible to determine sequences of a large number of clones for a short period of time.

It is unnecessary to determine the sequence for an entire length of the insert. It is enough to determine a length of nucleotide sequence which is considered to be sufficient to perform homology search. For example, in Examples described later on, the homology search as described below was performed when a sequence having not less than 60 nucleotides was successfully determined.

<3> Homology search with gene data base

The determined nucleotide sequence of each of cDNA clones is used to perform the homology search with respect to known amino acid sequences of the cellulose synthase or nucleotide sequences of genes coding therefor registered in the gene data base. The cellulose synthase is exemplified by an enzyme encoded by a gene of cellulose 4- $\beta$ -glucosyltransferase (BcsA) of cellulose synthase operon of acetobacterium (Wong, H. C. et al., Proc. Natl. Acad. Sci. U.S.A., 87, 8130-8134 (1990), ACCESSION No. M37202).

Those usable as the data base include, for example, GenBank, EMBL, and DDBJ published, for example, from Los Alamos National Institute in the United States, Institute of European Molecular Biology, and National Institute of Genetics (Japan). Those commercially available and useable as the program for homology search include, for example, commercially available DNA analysis softwares, such as DNASIS (Hitachi Software Engineering Co., Ltd.) and GENE-TYX (SDC Software Development). The following methods are also available. That is, a computer terminal is connected with the host computer of National Institute of Genetics to perform analysis. Alternatively, a personal computer is connected on Internet with NCBI (National Center for Biotechnology Information) to utilize (<http://www.ncbi.nlm.nih.gov/BLAST/>) BLAST (Basic Local Alignment Search Tool) so that high speed homology search is performed.

The homology search is performed, for example, in accordance with the following algorithm. When the homology search is performed for a nucleotide sequence, homology comparison is advanced while shifting the nucleotide sequence to be investigated by every one nucleotide with respect to individual gene sequences included in the data base. When six or more continuous nucleotides are coincident, the homology score is counted and calculated in accordance with a homology score table (see, for example, M. Dayhoff, Atlas of Protein Sequence and Structure, vol. 5 (1978)). The system is set so that those having a score not less than a certain value are picked up as candidates which have homology. Further, the gap may be introduced into the sequence to be investigated or into the gene sequence included in the data base to make optimization so that the score is maximized.

When the homology search is performed for an amino acid sequence, a nucleotide sequence to be investigated is converted into amino acids concerning all six frames including those of a complementary chain. The investigation may be performed in the same manner as performed for the nucleotide. Specifically, it is possible to use blastx of BLAST described above. As for detailed techniques and conditions for the search, reference may be made to DDBJ News Letter, No. 15 (February 1995).

<4> Isolation of cDNA clone of cotton plant cellulose synthase

The clone obtained as described above is not necessarily contain the entire nucleotide sequence of the gene. In such a case, the clone is used as a probe to perform screening by means of plaque hybridization. Thus, it is possible to obtain a clone containing a full length gene from the library. A specified method may be carried out with reference to Molecular Cloning, second edition (Maniatis et al., Cold Spring Harbour Laboratory) 12.30 to 12.40.

When obtained cDNA is deficient in 5'-portion, the 5'-portion can be obtained as well by synthesizing primers so that the cDNA sequence may be elongated toward the 5'-terminal, and performing RT-PCR by using mRNA as a template.

As demonstrated in Examples described later on, the DNA of the present invention has been obtained as those having homology to the known bacterial cellulose synthase gene. The DNA further codes for an amino acid sequence GlnXXXXXXArgTrp (SEQ ID NO: 12) which is considered to form a UDP-glucose binding domain, having high homology in the vicinity thereof.

The nucleotide sequences of DNA of the present invention obtained as described above and the amino acid sequences deduced from the nucleotide sequences are shown in SEQ ID NOs: 1 to 10 in Sequence Listing. SEQ ID NOs: 1 and 3 show nucleotide sequences of PcsA1 and PcsA2 respectively. SEQ ID NOs: 2 and 4 show amino acid sequences deduced from the nucleotide sequences of PcsA1 and PcsA2 respectively.

SEQ ID NOs: 5 and 6 show a nucleotide sequence of a clone (PcsA3-682) containing 3'-side region of PcsA3 and an amino acid sequence deduced from the nucleotide sequence respectively. SEQ ID NOs: 7 and 8 show a nucleotide sequence of a 5'-portion (PcsA3-5') of another clone containing 5'-side region of PcsA3 and an amino acid sequence deduced from the nucleotide sequence respectively. SEQ ID NOs: 9 and 10 show a nucleotide sequence of 3'-portion (PcsA3-3') of the clone and an amino acid sequence deduced from the nucleotide sequence respectively (see Fig. 1). That is, SEQ ID NO: 5 corresponds to the 3'-side region of PcsA3, SEQ ID NO: 7 corresponds to the 5'-side region of PcsA3, and SEQ ID NO: 9 corresponds to internal region of PcsA3. The overlapping portion of PcsA3-682 is different from that of PcsA3-3' in 9 nucleotides in the nucleotide sequence and 1 amino acid in the amino acid sequence. Figs. 3 and 4 show the comparison between the nucleotide sequences of PcsA3-682 and PcsA3-3'. SEQ ID NO: 11 shows a combination of the amino acid sequences encoded by PcsA3-682 and PcsA3-3'.

The sequence of GlnXXXXXXArgTrp (SEQ ID NO: 12) corresponds to amino acid numbers 710 to 714 in SEQ ID NO: 2 for PcsA1, amino acid numbers 778 to 782 in SEQ ID NO: 4 for PcsA2, and amino acid numbers 356 to 360 in

SEQ ID NO: 6 for PcsA3.

PcsA1 is different from CelA1 reported by Pear et al. (Proceeding of National Academy of Science, USA (1996), 93, 12637-12642) in nucleotide sequence by 28 nucleotides. As a result, the former is different from the latter in amino acid sequence encoded thereby by 10 amino acid residues. In general, the sugar chain specificity and the substrate specificity of the sugar chain transferase are extremely changed by point mutation of the nucleotide of DNA (Yamamoto and Hakomori, The Journal of Biological Chemistry (1990) 265, 19257-19262). Therefore, it is unclear whether or not CelA1 codes for a protein having the cellulose synthase activity. Incidentally, the 48th Arg, the 56th Ser, the 81st Asn, the 104th Ala, the 110th Ser, the 247th Asp, the 376th Asp, the 386th Ser, the 409th Arg, and the 649th Ser in the amino acid sequence encoded by CelA1 correspond to Gln, Ile, Ser, Thr, Pro, Asn, Glu, Pro, His, and Gly in PcsA1 respectively.

PcsA2 of the present invention contains the same sequence as that of CelA2 reported by Pear et al. However, CelA2 has an incomplete length, and it does not contain the entire coding region. CelA2 corresponds to nucleotide numbers of 1083 to 3311 in the nucleotide sequence of PcsA2 shown in SEQ ID NO: 3.

Any of the amino acid sequences shown in SEQ ID NOs: 2, 4, 6, 8, 10, and 11 is a novel sequence. All genes having nucleotide sequences coding for the amino acid sequences are included in the present invention.

The amino acid sequences described above may include deletion, substitution, insertion, and/or addition of one or more amino acid residues provided that the characteristic of the gene of the present invention is not substantially affected. The deletion, substitution, insertion, and/or addition of one or more amino acid residues as described above is obtainable by modifying the DNA's coding for the amino acid sequences shown in SEQ ID NOs: 2, 4, 6, 8, 10, and 11 randomly in accordance with the ordinary mutation treatment or intentionally in accordance with the site-directed mutagenesis method. As described above, in general, the sugar chain specificity and the substrate specificity of the sugar chain transferase are extremely changed by point mutation of the nucleotide of DNA. Therefore, DNA coding for a protein having the cellulose synthase activity is selected from the modified DNA's. The cellulose synthase activity can be measured, for example, by means of the method described by T. Hayashi: Measuring- $\beta$ -glucan deposition in plant cell walls, in Modern Methods of Plant Analysis: Plant Fibers, eds. H. F. Linskens and J. F. Jackson, Springer-Verlag, 10: 138-160 (1989).

Those harboring proteins or genes partially different from the sequences shown in Sequence Listing may exist depending on, for example, the variety of cotton plant or natural mutation. However, such genes are also included in the gene of the present invention. Such a gene may be obtained as DNA which is hybridizable under the stringent condition with all or a part of the coding region of the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, or 9. The "stringent condition" referred to herein indicates a condition under which a so-called specific hybrid is formed, and non-specific hybrid is not formed. It is difficult to definitely express such a condition by using a numerical value. However, for example, the stringent condition is exemplified by a condition under which nucleic acids having high homology, for example, DNA's having homology of not less than 80 % undergo hybridization with each other, and nucleic acids having homology lower than the above do not undergo hybridization with each other.

#### <5> Utilization of gene of the present invention

The DNA of the present invention makes it possible to control the cellulose synthesis in prokaryotic cells such as acetobacterium and/or eukaryotic cells such as yeasts belonging to, for example, the genus Saccharomyces, cells of plant such as cotton plant, and cultured cells of mammals and the like.

Specifically, the cellulose synthesis in the cells as described above can be facilitated, for example, by connecting a promoter to an upstream region of the DNA of the present invention, inserting an obtained fragment into an appropriate vector to construct a recombinant vector, and introducing the vector into the cells. Alternatively, the cellulose synthesis in the cells can be suppressed by introducing an antisense gene of the DNA of the present invention into the cells.

The promoter and the vector may be selected from those ordinarily utilized to express heterogeneous genes, and the method ordinarily employed to express heterogeneous genes may be used as the transformation method. Specifically, in the case of yeast, it is possible to use a protein-expressing kit produced by Invitrogen, i.e., Pichia Expression Kit, and a vector pPIC9 contained in this kit. For example, COS7 cells may be used as mammalian cultured cells, and a vector CDM8 may be used therefor.

The present invention provides the DNA coding for cellulose synthase. The DNA provides a new method for controlling cellulose production by incorporating the DNA into prokaryotic cells and eukaryotic cells.

#### Brief Description of the Drawings

Fig. 1 shows a relationship between two clones of PcsA3 as an embodiment of the DNA of the present invention. Regions interposed between arrows indicate regions for which nucleotide sequences have been determined. A dotted line indicates a region for which no nucleotide sequence has been determined.

Fig. 2 shows a structure of *EcoRI* adapter.

Fig. 3 shows comparison between sequences of PcsA3-682 and PcsA3-3' (former half).

Fig. 4 shows comparison between sequences of PcsA3-682 and PcsA3-3' (latter half). "." indicates coincident nucleotides, and "\*" indicates non-coincident nucleotides.

## Best Mode for Carrying Out the Invention

Examples of the present invention will be explained below.

### <1> Preparation of total RNA from cotton plant

Cotton plant (*Gossypium hirsutum* L.) Coker 312 was used as a material. Fiber cells on 16 to 18 days post anthesis were collected in liquid nitrogen. The cotton plant fiber cells in an amount of 75 g were sufficiently ground in a mortar while being frozen with liquid nitrogen. Powdered fiber was transferred to a centrifuge tube equipped with a cap, to which 375 mg of DTT as a powder was added, followed by addition of 200 ml of XT buffer (obtained by adjusting 0.2 M sodium borate containing 30 mM EDTA and 1 % SDS to be pH 9.0, and then applying a diethylpyrocarbonate treatment, followed by autoclaving to obtain a solution to which vanadylribonucleoside was added to give a concentration of 10 mM) having been heated to 90 to 95 °C. An obtained solution was sufficiently agitated.

The solution was added with 100 mg of protease K, and it was agitated again. The solution was incubated at 40 °C for 2 hours, and then it was added with 16 ml of 2 M KCl. The solution was sufficiently agitated again, and it was left to stationarily stand in ice for 1 hour, followed by centrifugation for 20 minutes (4 °C) at 12,000 g by using a high speed refrigerated centrifuge.

An obtained supernatant was filtrated, and floating matters were removed. The solution was transferred to a measuring cylinder to measure the volume. The solution was transferred to another centrifuge tube, to which lithium chloride was added in an amount of 85 mg per 1 ml of the extract solution to give a final concentration of 2 M. The solution was left to stationarily stand at 4 °C overnight, and then precipitated RNA was separated by centrifugation for 20 minutes at 12,000 g. An obtained precipitate of RNA was washed and precipitated twice with cooled 2 M lithium chloride.

The obtained RNA was dissolved in 10 mM Tris buffer (pH 7.5) to give a concentration of about 2 mg/ml, to which 5 M potassium acetate was added to give a concentration of 200 mM. Ethanol was added thereto to give a concentration of 70 %, followed by cooling at -80 °C for 10 minutes. Centrifugation was performed at 4 °C for 10 minutes at 15,000 rpm, and then an obtained precipitate was suspended in an appropriate amount of sterilized water to give an RNA sample. As a result of quantitative measurement for the RNA sample, total RNA was obtained in an amount of 2 mg.

### <2> Purification of mRNA

mRNA was purified as a poly(A)<sup>+</sup> RNA fraction from the total RNA obtained as described above. Purification was performed by using Oligotex-dT30 <Super> (purchased from Toyobo) as oligo(dT)-immobilized latex for poly(A)<sup>+</sup> RNA purification.

Elution buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1 % SDS) was added to a solution containing 1 mg of the total RNA to give a total volume of 1 ml, to which 1 ml of Oligotex-dT30 <Super> was added, followed by heating at 65 °C for 5 minutes and quick cooling on ice for 3 minutes. The obtained solution was added with 0.2 ml of 5 M NaCl, and it was incubated at 37 °C for 10 minutes, followed by centrifugation at 15,000 rpm for 3 minutes. After that, a supernatant was carefully removed.

An obtained pellet was suspended in 2.5 ml of Washing Buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.5 M NaCl, 0.1 % SDS), and the suspension was centrifuged at 15,000 rpm for 3 minutes. After that, a supernatant was carefully removed. An obtained pellet was suspended in 1 ml of TE Buffer, and then it was heated at 65 °C 5 minutes. The suspension was quickly cooled on ice for 3 minutes, and then it was centrifuged at 15,000 rpm for 3 minutes to recover poly(A)<sup>+</sup> mRNA contained in an obtained supernatant.

Thus, the poly(A)<sup>+</sup> mRNA in an amount of about 10 µg was obtained from 1 mg of the total RNA. An aliquot of 5 µg thereof was used to prepare a cDNA library.

### <3> Preparation of cDNA library

#### (1) Synthesis of cDNA

The mRNA obtained as described above was used as a template to synthesis cDNA by using a λZAP cDNA synthesis kit produced by Stratagene. The following solution was prepared and mixed in a tube.

5.0 µl 10 x 1st Strand Buffer (buffer for reverse transcription reaction);  
 3.0 µl 10 mM 1st Strand Methyl Nucleotide Mix (5-methyl dCTP, dATP, dGTP, dTTP mixture);  
 2.0 µl Linker-Primer (linker and primer);  
 H<sub>2</sub>O (adjusted to give a total volume of 50 µl);  
 1.0 µl RNase Block II (RNase inhibitor).

The respective components described above were contents of the kit. Linker-Primer had a sequence as shown in SEQ ID NO: 13. Methylated nucleotide was used because it was intended not to allow cDNA to be digested by the restriction enzyme reaction performed later on. The reaction solution was agitated well, and then 5.0 µg of poly(A)<sup>+</sup> mRNA was added thereto, followed by being left to stand at room temperature for 10 minutes. Further, 2.5 µl of M-MuLV RTase (reverse transcriptase) was added (at this time, the total volume was 50 µl). The reaction solution was gently mixed, followed by centrifugation under a mild condition to allow the reaction solution to fall to the bottom of the tube. The reaction was performed at 37 °C for 60 minutes.

Next, the following solution was prepared and mixed in the tube in a certain order.

45.0 µl reaction solution containing cDNA primary chain;  
 40.0 µl 10 x 2nd Strand Buffer (buffer for polymerase reaction);  
 6.0 µl 2nd Strand Nucleotide Mixture (A, G, C, T mixture);  
 302.0 µl H<sub>2</sub>O.

The following solution was further added. However, in order to allow RNase and DNA polymerase to simultaneously act, enzyme solutions were allowed to adhere to the wall of the tube. After that, a vortex treatment was promptly performed, and the reaction solutions were allowed to fall to the bottom of the tube by means of centrifugation to perform a reaction for synthesizing cDNA second strand at 16 °C for 150 minutes.

0.8 µl RNase H (RNA-degrading enzyme);  
 7.5 µl DNA polymerase I (10.0 u/µl).

The reaction solution was added with 400 µl of a mixed solution of phenol: chloroform (1:1). Agitation was performed well, followed by centrifugation at room temperature for 2 minutes. An obtained supernatant was added with 400 µl of phenol: chloroform again, which was subjected to a vortex treatment and centrifugation at room temperature for 2 minutes. An obtained supernatant was added with the following solution to precipitate cDNA.

33.3 µl 3 M sodium acetate solution;  
 867.0 µl 100 % ethanol.

The obtained solution was left to stand at -20 °C overnight, and it was centrifuged at room temperature for 60 minutes. After that, washing was gently performed with 80 % ethanol, followed by centrifugation for 2 minutes. A supernatant was removed. An obtained pellet was dried, and it was dissolved in 43.5 µl of sterilized water. An aliquot (39.0 µl) was added with the following solution to blunt-end cDNA terminals.

5.0 µl 10 x T4 DNA Polymerase Buffer (buffer for T4 polymerase reaction);  
 2.5 µl 2.5 mM dNTP Mix (A, G, C, T mixture);  
 3.5 µl T4 DNA polymerase (2.9 u/µl).

The reaction was performed at 37 °C for 30 minutes, to which 50 µl of distilled water was added, and then 100 µl of phenol: chloroform was added thereto, followed by a vortex treatment and centrifugation for 2 minutes. An obtained supernatant was added with 100 µl of chloroform, which was subjected to a vortex treatment, followed by centrifugation for 2 minutes. The supernatant was added with the following solution to precipitate cDNA.

7.0 µl 3 M sodium acetate solution;  
 226 µl 100 % ethanol.

The solution was left to stand on ice for 30 minutes or more, and it was centrifuged at 4 °C for 60 minutes. An obtained precipitate was washed with 150 µl of 80 % ethanol, followed by centrifugation for 2 minutes and drying. The cDNA pellet was dissolved in 7.0 µl of EcoRI Adaptor solution, to which the following solution was added to ligate the EcoRI adaptor to both ends of the cDNA. Sequences of respective strands of the EcoRI adaptor are shown in SEQ ID NO: 14 and Fig. 2.

1.0 µl 10 x Ligation Buffer (buffer for ligase reaction);  
 1.0 µl 10 mM ATP;  
 1.0 µl T4 DNA ligase.

5 The reaction solution was centrifuged under a mild condition, and it was left to stand at 4 °C overnight or more. The solution was treated at 70 °C for 30 minutes, and then it was centrifuged under a mild condition, followed by being left to stand at room temperature for 5 minutes. The reaction solution was added with the following solution to phosphorylate 5'-terminals of the EcoRI adapter.

10 1.0 µl 10 x Ligation Buffer (buffer for ligase reaction);  
 2.0 µl 10 mM ATP;  
 6.0 µl H<sub>2</sub>O;  
 1.0 µl T4 polynucleotide kinase (10.0 u/µl).

15 The reaction was performed at 37 °C for 30 minutes, followed by a treatment at 70 °C for 30 minutes. The solution was centrifuged under a mild condition, and it was left to stand at room temperature for 5 minutes. The following solution was further added thereto to perform a reaction at 37 °C for 90 minutes so that the XhoI site introduced by Linker-Primer was digested with XhoI, followed by being left to stand at room temperature to perform cooling.

20 28.0 µl XhoI Buffer;  
 3.0 µl XhoI (45 u/µl).

The reaction solution was added with 5.0 µl of 10 x STE (10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA), which was added into a centrifuge column for removing short fragments (Sephacryl Spin Column) to perform centrifugation at 600 g for 2 minutes to obtain an eluent which was designated as Fraction 1. This operation was further repeated three times to obtain Fractions 2, 3, and 4 respectively. Fractions 3 and 4 were combined, to which phenol:chloroform (1:1) was added and agitated well, followed by centrifugation at room temperature for 2 minutes. An obtained supernatant was added with an equal amount of chloroform, and an obtained mixture was agitated well. The mixture was centrifuged at room temperature for 2 minutes to obtain a supernatant to which a two-fold amount of 100 % ethanol was added, followed by being left to stand at - 20 °C overnight. The solution was centrifuged at 4 °C for 60 minutes, followed by washing with an equal amount of 80 % ethanol. The solution was centrifuged at 4 °C for 60 minutes to obtain a cDNA pellet which was suspended in 10 µl of sterilized water.

## 35 (2) Preparation of cDNA library

The double strand cDNA obtained as described above was ligated with λ phage expression vector to prepare a recombinant vector. The following solution was prepared and mixed in a tube to perform a reaction at 12 °C overnight, followed by being left to stand at room temperature for 2 hours to ligate cDNA with the vector.

40 2.5 µl cDNA solution;  
 0.5 µl 10 x Ligation Buffer;  
 0.5 µl 10 mM ATP;  
 1.0 µl λZAP vector DNA (1 µg/µl);  
 0.5 µl T4 DNA ligase (4 Weiss u/µl).

## 45 (3) Packaging of phage DNA into phage particles

The phage vector containing the cDNA was packaged into phage particles by using an *in vitro* packaging kit (Gigapack II Gold packaging extract: produced by Stratagene). The recombinant phage solution was added to Freeze/Thaw extract immediately after dissolution, and the solution was placed on ice, to which 15 µl of Sonic extract was added to perform mixing well by pipetting. The reaction solution was centrifuged under a mild condition, and it was left to stand at room temperature (22 °C) for 2 hours. The reaction solution was added with 500 µl of Phage Dilution Buffer, to which 20 µl of chloroform was further added, followed by mixing. In order to measure the titer of the library, an aliquot (2 µl) of 500 µl of the aqueous phase was diluted in a ratio of 1:10 with 18 µl of SM buffer (5.8 g of NaCl, 2 g of MgSO<sub>4</sub>·7H<sub>2</sub>O, 50 ml of 1 M Tris-HCl (pH 7.5), and 5 ml of 2 % gelatin in 1 L). The diluted solution (1 µl) and the phage stock solution (1 µl) were plated respectively together with 200 p1 of a culture solution of *Escherichia coli* PLK-F' strain having been cultivated to arrive at a value of OD<sub>600</sub> of 0.5. That is, *Escherichia coli* PLK-F' strain was mixed with the phage solution to perform cultivation at 37 °C for 15 minutes. The obtained culture was added to 2 to 3 ml of top agar



(48 °C), which was immediately overlaid on NZY agar plate having been warmed at 37 °C. Cultivation was performed overnight at 37 °C, and appeared plaques were counted to calculate the titer. As a result, the titer was  $1.2 \times 10^6$  pfu/ml.

#### (4) Amplification of library

A centrifuge tube was added with the packaging solution containing about 50,000 recombinant bacteriophages and 600 µl of a culture solution of *Escherichia coli* PLK-F<sup>+</sup> strain having been cultivated to have a value of OD<sub>600</sub> of 0.5, followed by cultivation at 37 °C for 15 minutes. The culture solution was added with 6.5 ml of top agar having been maintained at 48 °C after dissolution, which was overlaid on 150 mm NZY plate having been warmed at about 37 °C, followed by cultivation at 37 °C for 5 to 8 hours. The respective plates were added with 10 ml of SM Buffer to perform cultivation at 4 °C overnight with gentle shaking. SM Buffer in the respective plates was collected in a sterilized polypropylene tube. The respective plates were rinsed with 2 ml of SM Buffer, and the rinsing solutions were collected in the same tube. Chloroform in an amount corresponding to 5 % of the total amount was added and mixed, followed by being left to stand at room temperature for 15 minutes. Bacterial cells were removed by centrifugation at 4,000 g for 5 minutes. An obtained supernatant was added with chloroform in an amount corresponding to 0.3 % of the total amount, and it was stored at 4 °C. The titer of the library amplified as described above was measured in the same manner as described above. As a result, the titer was  $2.3 \times 10^9$  pfu/ml.

#### (5) Excision of plasmid from phage DNA

*In vivo* excision of the plasmid portion from the recombinant phage DNA was performed. The following solution was mixed in 50 ml of a conical tube to cause infection at 37 °C for 15 minutes:

culture solution of *Escherichia coli* XL1-Blue (OD<sub>600</sub> = 0.1) 200 µl;  
phage solution after amplification 200 µl ( $> 1 \times 10^5$  phage particles);  
helper phage F408 1 µl ( $> 1 \times 10^6$  pfu/ml).

The mixed solution was added with 5 ml of 2 x YT medium to perform cultivation at 37 °C for 3 hours with shaking. A heat treatment was applied thereto at 70 °C for 20 minutes, followed by centrifugation at 4,000 g for 5 minutes. An obtained supernatant was decanted and transferred to a sterilized tube. Centrifugation was performed to obtain a supernatant which was diluted 100 times to obtain a solution. An aliquot (20 µl) of the solution was mixed with 200 µl of a culture solution of *Escherichia coli* XL1-Blue having been cultivated to obtain a value of OD<sub>600</sub> of 1.0 to cause infection at 37 °C for 15 minutes. Aliquots (1 to 100 µl) of the culture solution were plated on LB plates containing ampicillin, followed by cultivation at 37 °C overnight. Appeared colonies were randomly selected. Selected colonies were added with glycerol, and they were stored at -80 °C.

#### (6) Preparation of plasmid

Plasmids were prepared by using Magic Mini-prep kit produced by Promega. The culture fluid of *Escherichia coli* harboring the plasmid having been stored at -80 °C was inoculated into 5 ml of 2 x YT medium, followed by cultivation at 37 °C overnight. Centrifugation was performed for 5 minutes (4,000 rpm, 4 °C), and a supernatant was removed by decantation. An obtained bacterial cell pellet was added with 1 ml of TE buffer, followed by a vortex treatment. An obtained bacterial cell suspension was transferred to an Eppendorf tube, followed by centrifugation for 5 minutes (5,000 rpm, 4 °C). A resultant supernatant was removed by decantation.

An obtained bacterial cell pellet was added with 300 µl of Cell Resuspension Solution, and it was sufficiently suspended therein. An obtained suspension was transferred to an Eppendorf tube. The suspension was agitated for 2 minutes with a mixer, to which 300 µl of Cell Lysis Solution was added, followed by agitation until the suspension became transparent. Neutralization Solution (300 µl) was added thereto, and agitation was performed by shaking with the hand, followed by centrifugation for 10 minutes (15,000 rpm).

Only an obtained supernatant was transferred to a new Eppendorf tube (1.5 ml). A suction tube was prepared, to which a cock, a miniature column and a syringe (injector) were connected in this order. A resin in an amount of 1 ml was charged into the syringe. The supernatant was poured into the syringe, and agitation was performed well, followed by suction. Column Washing Solution in an amount of 2 ml was added, and washing was performed while performing suction. Suction was continued for 1 to 2 minutes in order to dry up. The miniature column was removed from the equipment, and it was set in a new Eppendorf tube (1.5 ml). Sterilized water in an amount of 100 µl having been warmed at 65 to 70 °C was poured into the miniature column, and the column and the Eppendorf tube were centrifuged together for 1 minute (5,000 rpm). An eluted solution was transferred to an Eppendorf tube, to which 5 µl of 3 M sodium acetate aqueous solution was added, and 250 µl of cold ethanol was added thereto. The solution was centrifuged (15,000 rpm,

25 minutes), and a supernatant was discarded. An obtained precipitate was added with 1 ml of 70 % ethanol, followed by centrifugation again (15,000 rpm, 3 minutes). Ethanol was completely removed, and the tube was vacuum-dried in a desiccator. The precipitate was sufficiently dissolved in 20 µl of sterilized water, and an obtained solution was stored at -20 °C. An aliquot (1 µl) of the solution was dispensed, and it was subjected to electrophoresis together with volume markers to quantitatively determine the plasmid DNA.

#### <4> Determination of nucleotide sequence of cDNA and homology search with gene data base

##### (1) Determination of nucleotide sequence of cDNA

The nucleotide sequence of cDNA was analyzed by using DNA automatic sequencer 373A produced by Applied Biosystems Inc. (ABI). The sequencing reaction was performed in accordance with an attached manual by using T3 primer based on the use of Dye Primer Cycle Sequencing Kit produced by the same company. The nucleotide sequence was determined for about 750 clones which were randomly selected.

##### (2) Homology search

Partial sequences of about 750 clones were searched with a computer using BlastX. As a result, three clones appeared to be homologues of bacterial cellulose synthase subunit. Therefore, it was tried to isolate full length clones.

#### <5> Isolation of full length clones

##### (1) 5'-RACE

As a result of the homology search, the obtained homologue clones were found to be partial length clones. Therefore, primers were synthesized to make elongation toward the 5' upstream so that RT-PCR was performed by using mRNA as a template.

##### (1-a) Synthesis of first-strand DNA

The following solution was prepared and mixed in a tube.

0.5 µl 10 µmol gene-specific primer 1;  
1 pg total RNA;  
DEPC-treated H<sub>2</sub>O (adjusted to give a total amount of 9 µl).

The following oligonucleotides were used as the gene-specific primer 1. That is, an oligonucleotide having a nucleotide sequence shown in SEQ ID NO: 15 was used for PcsA1. An oligonucleotide having a nucleotide sequence shown in SEQ ID NO: 16 was used for PcsA2. An oligonucleotide having a nucleotide sequence shown in SEQ ID NO: 17 was used for PcsA3.

The reaction solution was gently mixed, and then it was centrifuged under a mild condition to allow the reaction solution to fall to the bottom of the tube. The solution was left to stand at 70 °C for 10 minutes, followed by immediate cooling on ice.

Next, the following solution was prepared and mixed in the tube.

5 x RT Buffer 5 p1;  
25 mM MgCl<sub>2</sub> 2.5 µl;  
2 mM dNTP mix 5 µl;  
0.1 M DTT 2.5 µl;  
H<sub>2</sub>O (added to give a total amount of 24 µl).

The solution was gently agitated, and then it was centrifuged under a mild condition to allow the reaction solution to fall to the bottom of the tube, followed by being left to stand at 42 °C for 1 minute. The solution was added with 1 µl of SuperScriptIII RT (reverse transcriptase, GIBCO BRL), and it was gently mixed. After that, the reaction was performed at 42 °C for 50 minutes. Subsequently, the reaction solution was left to stand at 70 °C for 15 minutes to stop the reaction. Centrifugation was performed under a mild condition to allow the reaction solution to fall to the bottom of the tube, followed by being left to stand at 37 °C. RNase H (produced by Toyobo) in an amount of 1 µl was added thereto to perform a reaction at 37 °C for 30 minutes.

Subsequently, in order to remove excessive primers and nucleotides contained in the reaction solution, gel filtration was performed by using a purification column produced by Boehringer, Quick Spin Columns. At first, the tip of the column was removed, followed by centrifugation at 1,100 x g for 2 minutes to discard the buffer. The reaction solution was introduced into the central area of the column, followed by centrifugation at 1,100 x g for 4 minutes to recover the solution.

#### (1-b) Poly(dC) tailing

An aliquot (5 µl) was dispensed from the obtained solution, to which the following solution was added.

5 µl 5 x CoCl<sub>2</sub> Buffer;  
2.5 µl 2 mM dCTP;  
H<sub>2</sub>O (adjusted to give a total amount of 24 µl).

The reaction solution was mixed well, and it was left to stand at 94 °C for 3 minutes. Centrifugation was performed under a mild condition to allow the reaction solution to fall to the bottom of the tube, followed by being left to stand on ice. Terminal transferase TdT (produced by Toyobo) was added thereto in an amount of 1 µl, followed by mixing under a mild condition to perform a reaction at 37 °C for 10 minutes. Subsequently, the reaction solution was left to stand at 65 °C for 10 minutes to stop the reaction.

#### (1-c) PCR reaction

An aliquot (2.5 µl) was dispensed from the reaction solution, to which the following solution was added.

2.5 µl 10 x PCR Buffer;  
2.5 µl 2 mM dNTP mix;  
0.5 µl Gene-specific primer 2;  
0.5 µl Abridged Anchor Primer (GIBCO BRL);  
0.5 µl Advantage KlenTaq Polymerase Mix (Clontech);  
H<sub>2</sub>O (adjusted to give a total amount of 25 µl).

The following oligonucleotides were used as Gene-specific primer 2. That is, an oligonucleotide having a nucleotide sequence shown in SEQ ID NO: 18 was used for PcsA1. An oligonucleotide having a nucleotide sequence shown in SEQ ID NO: 19 was used for PcsA2. An oligonucleotide having a nucleotide sequence shown in SEQ ID NO: 20 was used for PcsA3.

The solution was introduced into a 0.2 ml tube to perform the PCR reaction under the following condition.

PAD	94 °C	90 seconds
30 cycles	94 °C	30 seconds
	60 to 68 °C	30 to 60 seconds
	68 °C	180 seconds
Final	68 °C	7 minutes
Hold	4 °C	

The reaction solution was subjected to agarose gel electrophoresis to extract, from the gel, DNA's corresponding to portions having the largest size (about 1.8 K for PcsA1, about 2 K for PcsA2, and about 2.2 K for PcsA3). GENO-BIND produced by CLONTECH was used for the extraction, and the procedure was carried out in accordance with its protocol. The DNA thus obtained was subjected to Poly(dC) tailing, which was used as a template to perform the PCR reaction. The condition and the composition of the reaction solution were the same as those described above.

## (2) Cloning

### (2-a) 5'-RACE TA cloning

Starting from the obtained PCR reaction solution, cloning was performed by using TA Cloning Kit produced by Invitrogen in accordance with its protocol.

The following solution was added to an aliquot (1.5 µl) of the PCR reaction solution obtained as described above.

0.5 µl 10 x Ligation Buffer;  
 1 µl pCR111 vector;  
 0.5 µl T4 DNA Ligase;  
 1.5 µl dH<sub>2</sub>O.

The reaction was performed at 14 °C overnight. An aliquot (2 p1) of the reaction solution was added to 25 p1 of *Escherichia coli* competent cell (JM109) preparation, followed by being left to stand for 30 minutes on ice. After that, heat shock was applied at 42 °C for 30 seconds. The solution was stationarily left to stand on ice for 2 minutes, to which 450 µl of SOB medium was thereafter added to perform cultivation at 37 °C for 1 hour with shaking at 200 rpm. The culture was spread over Amp/Xgal/IPTG plate, followed by incubation at 37 °C overnight. The plasmid was extracted from obtained colonies in accordance with the method as described above.

#### (2-b) Cloning of complete length cDNA

The procedure was carried out by using DNA Sequencer 377 produced by ABI in accordance with its protocol. The sequencing reaction was performed by using M13 primer and synthetic oligomer as primers, based on the use of Dye Terminator Cycle Sequencing Kit produced by the same company. As a result of the sequencing, as for PcsA3, it was revealed that another clone also belonging to the group of PcsA3 but having a slightly different sequence (one position for amino acid) was isolated (see Figs. 3 and 4). A nucleotide sequence of a clone (PcsA3-682) containing the 3'-side region of PcsA3 and an amino acid sequence deduced from this nucleotide sequence are shown in SEQ ID NOs: 5 and 6. A nucleotide sequence of a 5'-portion (PcsA3-5') of another clone containing the 5'-side region of PcsA3 and an amino acid sequence deduced from this nucleotide sequence are shown in SEQ ID NOs: 7 and 8. A nucleotide sequence of a 3'-portion (PcsA3-3') of the clone and an amino acid sequence deduced from this nucleotide sequence are shown in SEQ ID NOs: 9 and 10.

As for PcsA1 and PcsA2, primers for 5'-terminal and 3'-terminal of a region containing ORF were synthesized on the basis of the obtained sequences to perform the PCR reaction. Thus, complete length clones were isolated by means of TA cloning. The condition and the composition of the reaction solution were the same as those described above.

Oligonucleotides shown in SEQ ID NO: 21 (5'-terminal) and SEQ ID NO: 22 (3'-terminal) were used as the primers for PcsA1. Oligonucleotides shown in SEQ ID NO: 23 (5'-terminal) and SEQ ID NO: 24 (3'-terminal) were used as the primers for PcsA2. Results are shown in SEQ ID NOs: 1 to 4.

Ann x to th description

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: NISSHINBO INDUSTRIES, INC.

HAYASHI, Takahisa

(ii) TITLE OF INVENTION: CELLULOSE SYNTHASE GENE

(iii) NUMBER OF SEQUENCES: 24

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE:

(B) STREET:

(C) CITY:

(E) COUNTRY:

(F) ZIP:

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: JP 9-83133

(B) FILING DATE: 1-APR-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME:

(B) REGISTRATION NUMBER:

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE:

(B) TELEFAX:

## (2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Gossypium hirsutum L.

(C) INDIVIDUAL ISOLATE: Coker312

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 77..3001

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

	GGTTAGCATA TTGTTTGTAG CATTGGGTTT TTTTCTCAAG GAAGAAGAAG GAGAAAGATA	60
	AGTAATGTTT TTGAGA ATG ATG GAA TCT GGG GTT OCT GTT TGC CAC ACT	109
10	Met Met Glu Ser Gly Val Pro Val Cys His Thr	
	1 5 10	
	TGT GGT GAA CAT GTT GGG TTG AAT GTT AAT GGT GAA CCC TTT GTG GCT	157
	Cys Gly Glu His Val Gly Leu Asn Val Asn Gly Glu Pro Phe Val Ala	
15	15 20 25	
	TGC CAT GAA TGT AAT TTC OCT ATT TGT AAG AGT TGT TTT GAG TAT GAT	205
	Cys His Glu Cys Asn Phe Pro Ile Cys Lys Ser Cys Phe Glu Tyr Asp	
	30 35 40	
20	CTT AAG GAA GGA CAA AAA GCT TGC TTG OGT TGT GGT ATT COG TAT GAT	253
	Leu Lys Glu Gly Gln Lys Ala Cys Leu Arg Cys Gly Ile Pro Tyr Asp	
	45 50 55	
	GAA AAC CTG TTG GAC GAT GTC GAG AAG GCC ACC GGC GAT CAA TOG ACA	301
25	Glu Asn Leu Leu Asp Asp Val Glu Lys Ala Thr Gly Asp Gln Ser Thr	
	60 65 70 75	
	ATG GCT GCA CAT TTG AGC AAG TCT CAG GAT GTT GGA ATT CAT GCA AGA	349
	Met Ala Ala His Leu Ser Lys Ser Gln Asp Val Gly Ile His Ala Arg	
30	80 85 90	
	CAT ATC AGC AGT GTG TCT ACA TTG GAT AGT GAA ATG ACT GAA GAC AAT	397
	His Ile Ser Ser Val Ser Thr Leu Asp Ser Glu Met Thr Glu Asp Asn	
	95 100 105	
35	GGG AAT COG ATT TGG AAG AAC AGG GTG GAA AGT TGG AAA GAA AAG AAG	445
	Gly Asn Pro Ile Trp Lys Asn Arg Val Glu Ser Trp Lys Glu Lys Lys	
	110 115 120	
	AAC AAG AAG AAG AAG OCT GCA ACA ACT AAG GTT GAA AGA GAG GCT GAA	493
40	Asn Lys Lys Lys Lys Pro Ala Thr Thr Lys Val Glu Arg Glu Ala Glu	
	125 130 135	
	ATC OCA OCT GAG CAA CAA ATG GAA GAT AAA COG GCA COG GAT GCT TOC	541
	Ile Pro Pro Glu Gln Gln Met Glu Asp Lys Pro Ala Pro Asp Ala Ser	
45	140 145 150 155	
	CAG CCC CTC TOG ACT ATA ATT CCA ATC COG AAA AGC AGA CTT GCA CCA	589
	Gln Pro Leu Ser Thr Ile Ile Pro Ile Pro Lys Ser Arg Leu Ala Pro	
	160 165 170	
50	TAC OGA ACC GTG ATC ATT ATG OGA TTG ATC ATT CTC GGT CTT TTC TTC	637
	Tyr Arg Thr Val Ile Ile Met Arg Leu Ile Ile Leu Gly Leu Phe Phe	

55

				175				180					185				
	CAT	TAT	OGA	GTA	ACA	AAC	CCC	GTT	GAC	AGT	GCT	TTT	GGA	CTG	TGG	CTC	685
5	His	Tyr	Arg	Val	Thr	Asn	Pro	Val	Asp	Ser	Ala	Phe	Gly	Leu	Trp	Leu	
			190					195					200				
	ACT	TCA	GTC	ATA	TGT	GAA	ATC	TGG	TTT	GCT	TTT	TOC	TGG	GTG	TTG	GAT	733
	Thr	Ser	Val	Ile	Cys	Glu	Ile	Trp	Phe	Ala	Phe	Ser	Trp	Val	Leu	Asp	
			205					210					215				
10	CAG	TTC	OCT	AAG	TGG	TAT	OCT	GTT	AAC	AGG	GAA	ACA	TAC	ATT	GAC	AGA	781
	Gln	Phe	Pro	Lys	Trp	Tyr	Pro	Val	Asn	Arg	Glu	Thr	Tyr	Ile	Asp	Arg	
			220					225					230			235	
15	CTG	TCT	GCA	AGA	TAT	GAA	AGA	GAA	GGT	GAA	OCT	AAT	GAA	CTT	GCT	GCA	829
	Leu	Ser	Ala	Arg	Tyr	Glu	Arg	Glu	Gly	Glu	Pro	Asn	Glu	Leu	Ala	Ala	
				240									245			250	
	GTT	GAC	TTC	TTT	GTG	AGT	ACA	GTG	GAT	CCA	TTG	AAA	GAG	OCT	CCA	TTG	877
	Val	Asp	Phe	Phe	Val	Ser	Thr	Val	Asp	Pro	Leu	Lys	Glu	Pro	Pro	Leu	
20				255										260		265	
	ATT	ACT	GOC	AAT	ACT	GTG	CTT	TOC	ATC	CTT	GOC	TTG	GAC	TAC	COG	GTA	925
	Ile	Thr	Ala	Asn	Thr	Val	Leu	Ser	Ile	Leu	Ala	Leu	Asp	Tyr	Pro	Val	
				270										275		280	
25	GAT	AAG	GTC	TCT	TGT	TAT	ATA	TCT	GAT	GAT	GGT	GCG	GCC	ATG	CTG	ACA	973
	Asp	Lys	Val	Ser	Cys	Tyr	Ile	Ser	Asp	Asp	Gly	Ala	Ala	Met	Leu	Thr	
				285										290		295	
30	TTT	GAA	TCT	CTA	GTA	GAA	ACA	GCC	GAC	TTT	GCA	AGA	AAG	TGG	GTT	CCA	1021
	Phe	Glu	Ser	Leu	Val	Glu	Thr	Ala	Asp	Phe	Ala	Arg	Lys	Trp	Val	Pro	
				300										305		310	315
	TTC	TGC	AAA	AAA	TTT	TOC	ATT	GAA	CCA	CGG	GCA	OCT	GAG	TTT	TAC	TTC	1069
	Phe	Cys	Lys	Lys	Phe	Ser	Ile	Glu	Pro	Arg	Ala	Pro	Glu	Phe	Tyr	Phe	
35				320										325		330	
	TCA	CAG	AAG	ATT	GAT	TAC	TTG	AAA	GAT	AAA	GTG	CAG	CCC	TCT	TTT	GTA	1117
	Ser	Gln	Lys	Ile	Asp	Tyr	Leu	Lys	Asp	Lys	Val	Gln	Pro	Ser	Phe	Val	
				335										340		345	
40	AAA	GAA	CGT	AGA	GCT	ATG	AAA	AGA	GAT	TAC	GAA	GAG	TAC	AAA	ATT	CGA	1165
	Lys	Glu	Arg	Arg	Ala	Met	Lys	Arg	Asp	Tyr	Glu	Glu	Tyr	Lys	Ile	Arg	
				350										355		360	
45	ATC	AAT	GCT	TTA	GTT	GCA	AAG	GCT	CAG	AAA	ACA	OCT	GAA	GAA	GGA	TGG	1213
	Ile	Asn	Ala	Leu	Val	Ala	Lys	Ala	Gln	Lys	Thr	Pro	Glu	Glu	Gly	Trp	
				365										370		375	
	ACA	ATG	CAA	GAT	GGA	ACT	OCT	TGG	COG	GGA	AAT	AAC	COG	CGT	GAT	CAC	1261
	Thr	Met	Gln	Asp	Gly	Thr	Pro	Trp	Pro	Gly	Asn	Asn	Pro	Arg	Asp	His	
50				380				385					390			395	
	OCT	GGC	ATG	ATT	CAG	GTT	TTC	CTT	GGA	TAT	AGC	GGT	GCT	CAT	GAC	ATC	1309

55

	Pro Gly Met Ile Gln Val Phe Leu Gly Tyr Ser Gly Ala His Asp Ile	
	400 405 410	
5	GAA GGA AAT GAA CTT OCC OGA CTG GTT TAC GTC TCT AGA GAG AAG AGA	1357
	Glu Gly Asn Glu Leu Pro Arg Leu Val Tyr Val Ser Arg Glu Lys Arg	
	415 420 425	
	OCT GGC TAC CAA CAC CAC AAA AAG GCT GGT GCT GAA AAT GCT TTG GTT	1405
10	Pro Gly Tyr Gln His His Lys Lys Ala Gly Ala Glu Asn Ala Leu Val	
	430 435 440	
	AGG GTG TCT GCA GTT CTT ACA AAT GCT CCC TTC ATC CTC AAT CTT GAT	1453
	Arg Val Ser Ala Val Leu Thr Asn Ala Pro Phe Ile Leu Asn Leu Asp	
	445 450 455	
15	TGT GAC CAC TAT GTT AAC AAT AGC AAG GCA GTT AGG GAG GCA ATG TGC	1501
	Cys Asp His Tyr Val Asn Asn Ser Lys Ala Val Arg Glu Ala Met Cys	
	460 465 470 475	
	TTC TTG ATG GAC CCA CAA GTC GGT OGA GAT GTC TGC TAT GTG CAG TTT	1549
20	Phe Leu Met Asp Pro Gln Val Gly Arg Asp Val Cys Tyr Val Gln Phe	
	480 485 490	
	OCT CAA AGA TTT GAT GGC ATA GAT AGG AGT GAT OGA TAT GCC AAT OGG	1597
25	Pro Gln Arg Phe Asp Gly Ile Asp Arg Ser Asp Arg Tyr Ala Asn Arg	
	495 500 505	
	AAC ACA GTT TTC TTT GAT GTT AAC ATG AAA GGT CTT GAT GGA ATC CAA	1645
	Asn Thr Val Phe Phe Asp Val Asn Met Lys Gly Leu Asp Gly Ile Gln	
	510 515 520	
30	GGG OCT GTT TAT GTG GGA ACA GGT TGT GTT TTC AAT AGG CAA GCA CTT	1693
	Gly Pro Val Tyr Val Gly Thr Gly Cys Val Phe Asn Arg Gln Ala Leu	
	525 530 535	
	TAT GGC TAT GGT CCA OCT TCA ATG CCA AGT TTT CCC AAG TCA TOC TOC	1741
35	Tyr Gly Tyr Gly Pro Pro Ser Met Pro Ser Phe Pro Lys Ser Ser Ser	
	540 545 550 555	
	TCA TCT TGC TCG TGT TGC TGC CCC GGC AAG AAG GAA OCT AAA GAT CCA	1789
	Ser Ser Cys Ser Cys Cys Cys Pro Gly Lys Lys Glu Pro Lys Asp Pro	
40	560 565 570	
	TCA GAG CTT TAT AGG GAT GCA AAA CGG GAA GAA CTT GAT GCT GCC ATC	1837
	Ser Glu Leu Tyr Arg Asp Ala Lys Arg Glu Glu Leu Asp Ala Ala Ile	
	575 580 585	
45	TTT AAC CTT AGG GAA ATT GAC AAT TAT GAT GAG TAT GAA AGA TCA ATG	1885
	Phe Asn Leu Arg Glu Ile Asp Asn Tyr Asp Glu Tyr Glu Arg Ser Met	
	590 595 600	
	TTG ATC TCT CAA ACA AGC TTT GAG AAA ACT TTT GGC TTA TCT TCA GTC	1933
50	Leu Ile Ser Gln Thr Ser Phe Glu Lys Thr Phe Gly Leu Ser Ser Val	
	605 610 615	

55



	TTC	ATT	GAA	TCT	ACA	CTA	ATG	GAG	AAT	GGA	GTG	GCT	GAA	TCT	GCC	1981	
	Phe	Ile	Glu	Ser	Thr	Leu	Met	Glu	Asn	Gly	Gly	Val	Ala	Glu	Ser	Ala	
5	620					625				630					635		
	AAC	OCT	TOC	ACA	CTA	ATC	AAG	GAA	GCA	ATT	CAT	GTC	ATC	GGC	TGT	GGC	2029
	Asn	Pro	Ser	Thr	Leu	Ile	Lys	Glu	Ala	Ile	His	Val	Ile	Gly	Cys	Gly	
					640					645					650		
10	TAT	GAG	GAG	AAG	ACT	GCA	TGG	GGG	AAA	GAG	ATT	GGA	TGG	ATA	TAT	GGT	2077
	Tyr	Glu	Glu	Lys	Thr	Ala	Trp	Gly	Lys	Glu	Ile	Gly	Trp	Ile	Tyr	Gly	
					655					660					665		
	TCA	GTC	ACT	GAG	GAT	ATC	TTA	AOC	GGC	TTC	AAA	ATG	CAC	TGC	CGA	GGA	2125
15	Ser	Val	Thr	Glu	Asp	Ile	Leu	Thr	Gly	Phe	Lys	Met	His	Cys	Arg	Gly	
					670					675					680		
	TGG	AGA	TOG	ATT	TAC	TGC	ATG	CCC	TTA	AGG	OCA	GCA	TTC	AAA	GGA	TCT	2173
	Trp	Arg	Ser	Ile	Tyr	Cys	Met	Pro	Leu	Arg	Pro	Ala	Phe	Lys	Gly	Ser	
					685					690					695		
20	GCA	CCC	ATC	AAT	CTG	TCT	GAT	CGG	TTG	CAC	CAG	GTT	CTT	OGA	TGG	GCT	2221
	Ala	Pro	Ile	Asn	Leu	Ser	Asp	Arg	Leu	His	Gln	Val	Leu	Arg	Trp	Ala	
					700					705					710	715	
	CTT	GGA	TCT	GTT	GAA	ATT	TTC	CTA	AGC	AGG	CAT	TGC	OCT	CTA	TGG	TAT	2269
25	Leu	Gly	Ser	Val	Glu	Ile	Phe	Leu	Ser	Arg	His	Cys	Pro	Leu	Trp	Tyr	
					720					725					730		
	GGC	TTT	GGA	GGT	GGT	OGT	CTT	AAA	TGG	CTT	CAA	AGA	CTA	GCA	TAT	ATA	2317
	Gly	Phe	Gly	Gly	Gly	Arg	Leu	Lys	Trp	Leu	Gln	Arg	Leu	Ala	Tyr	Ile	
30					735					740					745		
	AAC	AOC	ATT	GTC	TAT	OCT	TTC	ACA	TOC	CTT	OCA	CTC	ATT	GOC	TAT	TGT	2365
	Asn	Thr	Ile	Val	Tyr	Pro	Phe	Thr	Ser	Leu	Pro	Leu	Ile	Ala	Tyr	Cys	
					750					755					760		
35	TCA	CTA	CCA	GCA	ATC	TGT	CTT	CTC	ACA	GGA	AAA	TTT	ATC	ATA	CCA	ACG	2413
	Ser	Leu	Pro	Ala	Ile	Cys	Leu	Leu	Thr	Gly	Lys	Phe	Ile	Ile	Pro	Thr	
					765					770					775		
	CTC	TCA	AAC	CTG	GCA	AGT	GTT	CTC	TTT	CTT	GGC	CTT	TTC	CTT	TOC	ATT	2461
40	Leu	Ser	Asn	Leu	Ala	Ser	Val	Leu	Phe	Leu	Gly	Leu	Phe	Leu	Ser	Ile	
					780					785					790	795	
	ATC	GTG	ACT	GCT	GTT	CTC	GAG	CTC	OGA	TGG	AGT	GGT	GTC	AGC	ATT	GAG	2509
	Ile	Val	Thr	Ala	Val	Leu	Glu	Leu	Arg	Trp	Ser	Gly	Val	Ser	Ile	Glu	
					800					805					810		
45	GAC	TTA	TGG	OGT	AAC	GAG	CAG	TTT	TGG	GTC	ATC	GGT	GGC	GTT	TCA	GOC	2557
	Asp	Leu	Trp	Arg	Asn	Glu	Gln	Phe	Trp	Val	Ile	Gly	Gly	Val	Ser	Ala	

	830	835	840	
	GAC ACC AAC TTT ACT GTC ACT GGC AAA GCA GCT GAT GAT GCA GAT TTT			2653
5	Asp Thr Asn Phe Thr Val Thr Ala Lys Ala Ala Asp Asp Ala Asp Phe			
	845	850	855	
	GGT GAG CTC TAC ATT GTG AAA TGG ACT ACA CTT CTA ATC OCT OCA ACA			2701
	Gly Glu Leu Tyr Ile Val Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr			
10	860	865	870	875
	ACA CTC CTC ATC GTC AAC ATG GTT GGT GTC GTT GGC GGA TTC TOC GAT			2749
	Thr Leu Leu Ile Val Asn Met Val Gly Val Val Ala Gly Phe Ser Asp			
	880	885	890	
15	GGC CTC AAC AAA GGG TAC GAA GCT TGG GGA OCA CTC TTT GGC AAA GTG			2797
	Ala Leu Asn Lys Gly Tyr Glu Ala Trp Gly Pro Leu Phe Gly Lys Val			
	895	900	905	
	TTC TTT TOC TTC TGG GTC ATC CTC CAT CTT TAT OCA TTC CTC AAA GGT			2845
20	Phe Phe Ser Phe Trp Val Ile Leu His Leu Tyr Pro Phe Leu Lys Gly			
	910	915	920	
	CTT ATG GGA CCG CAA AAC AGG ACA OCA ACC ATT GTT GTC CTT TGG TCA			2893
	Leu Met Gly Arg Gln Asn Arg Thr Pro Thr Ile Val Val Leu Trp Ser			
	925	930	935	
25	GTG TTG TTG GCT TCT GTC TTC TCT CTT GTT TGG GTT CGG ATC AAC CCG			2941
	Val Leu Leu Ala Ser Val Phe Ser Leu Val Trp Val Arg Ile Asn Pro			
	940	945	950	955
	TTT GTC AGC ACC GGC GAT AGC ACC ACC GTG TCA CAG AGC TGC ATT TOC			2989
30	Phe Val Ser Thr Ala Asp Ser Thr Thr Val Ser Gln Ser Cys Ile Ser			
	960	965	970	
	ATT GAT TGT TGATGATATT ATGTGTTTCT TAGAATTGAA ATCATTGCAA			3038
	Ile Asp Cys			
35	GTAAGTGGAC TGAAACATGT CTATTGACTA AGTTTGAAC AGTTTGTACC CATTTTATTC			3098
	TTAGCAGTGT GTAATTTTTC TAAACAATGC TATGAACTAT ACATATTTCA TTGATATTTA			3158
	CATTAAATGA AACTACATCA GTCTGCAGAA AAAAAAAAAA AAAAAAAAAA			3207

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 974 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met	Met	Glu	Ser	Gly	Val	Pro	Val	Cys	His	Thr	Cys	Gly	Glu	His	Val
1					5				10				15		
Gly	Leu	Asn	Val	Asn	Gly	Glu	Pro	Phe	Val	Ala	Cys	His	Glu	Cys	Asn

	20	25	30
	Phe Pro Ile Cys Lys Ser Cys Phe Glu Tyr Asp Leu Lys Glu Gly Gln		
5	35	40	45
	Lys Ala Cys Leu Arg Cys Gly Ile Pro Tyr Asp Glu Asn Leu Leu Asp		
	50	55	60
	Asp Val Glu Lys Ala Thr Gly Asp Gln Ser Thr Met Ala Ala His Leu		
10	65	70	75
	Ser Lys Ser Gln Asp Val Gly Ile His Ala Arg His Ile Ser Ser Val		
	85	90	95
	Ser Thr Leu Asp Ser Glu Met Thr Glu Asp Asn Gly Asn Pro Ile Trp		
15	100	105	110
	Lys Asn Arg Val Glu Ser Trp Lys Glu Lys Lys Asn Lys Lys Lys Lys		
	115	120	125
	Pro Ala Thr Thr Lys Val Glu Arg Glu Ala Glu Ile Pro Pro Glu Gln		
20	130	135	140
	Gln Met Glu Asp Lys Pro Ala Pro Asp Ala Ser Gln Pro Leu Ser Thr		
	145	150	155
	Ile Ile Pro Ile Pro Lys Ser Arg Leu Ala Pro Tyr Arg Thr Val Ile		
25	165	170	175
	Ile Met Arg Leu Ile Ile Leu Gly Leu Phe Phe His Tyr Arg Val Thr		
	180	185	190
	Asn Pro Val Asp Ser Ala Phe Gly Leu Trp Leu Thr Ser Val Ile Cys		
30	195	200	205
	Glu Ile Trp Phe Ala Phe Ser Trp Val Leu Asp Gln Phe Pro Lys Trp		
	210	215	220
	Tyr Pro Val Asn Arg Glu Thr Tyr Ile Asp Arg Leu Ser Ala Arg Tyr		
35	225	230	235
	Glu Arg Glu Gly Glu Pro Asn Glu Leu Ala Ala Val Asp Phe Phe Val		
	245	250	255
	Ser Thr Val Asp Pro Leu Lys Glu Pro Pro Leu Ile Thr Ala Asn Thr		
40	260	265	270
	Val Leu Ser Ile Leu Ala Leu Asp Tyr Pro Val Asp Lys Val Ser Cys		
	275	280	285
	Tyr Ile Ser Asp Asp Gly Ala Ala Met Leu Thr Phe Glu Ser Leu Val		
	290	295	300
	Glu Thr Ala Asp Phe Ala Arg Lys Trp Val Pro Phe Cys Lys Lys Phe		
50	305	310	315
	Ser Ile Glu Pro Arg Ala Pro Glu Phe Tyr Phe Ser Gln Lys Ile Asp		
	325	330	335
	Tyr Leu Lys Asp Lys Val Gln Pro Ser Phe Val Lys Glu Arg Arg Ala		
55	340	345	350

Met Lys Arg Asp Tyr Glu Glu Tyr Lys Ile Arg Ile Asn Ala Leu Val  
 355 360 365  
 5 Ala Lys Ala Gln Lys Thr Pro Glu Glu Gly Trp Thr Met Gln Asp Gly  
 370 375 380  
 Thr Pro Trp Pro Gly Asn Asn Pro Arg Asp His Pro Gly Met Ile Gln  
 385 390 395 400  
 10 Val Phe Leu Gly Tyr Ser Gly Ala His Asp Ile Glu Gly Asn Glu Leu  
 405 410 415  
 Pro Arg Leu Val Tyr Val Ser Arg Glu Lys Arg Pro Gly Tyr Gln His  
 420 425 430  
 15 His Lys Lys Ala Gly Ala Glu Asn Ala Leu Val Arg Val Ser Ala Val  
 435 440 445  
 Leu Thr Asn Ala Pro Phe Ile Leu Asn Leu Asp Cys Asp His Tyr Val  
 450 455 460  
 20 Asn Asn Ser Lys Ala Val Arg Glu Ala Met Cys Phe Leu Met Asp Pro  
 465 470 475 480  
 Gln Val Gly Arg Asp Val Cys Tyr Val Gln Phe Pro Gln Arg Phe Asp  
 485 490 495  
 25 Gly Ile Asp Arg Ser Asp Arg Tyr Ala Asn Arg Asn Thr Val Phe Phe  
 500 505 510  
 Asp Val Asn Met Lys Gly Leu Asp Gly Ile Gln Gly Pro Val Tyr Val  
 515 520 525  
 30 Gly Thr Gly Cys Val Phe Asn Arg Gln Ala Leu Tyr Gly Tyr Gly Pro  
 530 535 540  
 Pro Ser Met Pro Ser Phe Pro Lys Ser Ser Ser Ser Ser Cys Ser Cys  
 545 550 555 560  
 35 Cys Cys Pro Gly Lys Lys Glu Pro Lys Asp Pro Ser Glu Leu Tyr Arg  
 565 570 575  
 Asp Ala Lys Arg Glu Glu Leu Asp Ala Ala Ile Phe Asn Leu Arg Glu  
 580 585 590  
 40 Ile Asp Asn Tyr Asp Glu Tyr Glu Arg Ser Met Leu Ile Ser Gln Thr  
 595 600 605  
 Ser Phe Glu Lys Thr Phe Gly Leu Ser Ser Val Phe Ile Glu Ser Thr  
 610 615 620  
 45 Leu Met Glu Asn Gly Gly Val Ala Glu Ser Ala Asn Pro Ser Thr Leu  
 625 630 635 640  
 Ile Lys Glu Ala Ile His Val Ile Gly Cys Gly Tyr Glu Glu Lys Thr  
 645 650 655  
 50 Ala Trp Gly Lys Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp  
 660 665 670  
 55 Ile Leu Thr Gly Phe Lys Met His Cys Arg Gly Trp Arg Ser Ile Tyr

675                      680                      685  
 Cys Met Pro Leu Arg Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu  
 690                      695                      700  
 Ser Asp Arg Leu His Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu  
 705                      710                      715                      720  
 Ile Phe Leu Ser Arg His Cys Pro Leu Trp Tyr Gly Phe Gly Gly Gly  
 725                      730                      735  
 Arg Leu Lys Trp Leu Gln Arg Leu Ala Tyr Ile Asn Thr Ile Val Tyr  
 740                      745                      750  
 Pro Phe Thr Ser Leu Pro Leu Ile Ala Tyr Cys Ser Leu Pro Ala Ile  
 755                      760                      765  
 Cys Leu Leu Thr Gly Lys Phe Ile Ile Pro Thr Leu Ser Asn Leu Ala  
 770                      775                      780  
 Ser Val Leu Phe Leu Gly Leu Phe Leu Ser Ile Ile Val Thr Ala Val  
 785                      790                      795                      800  
 Leu Glu Leu Arg Trp Ser Gly Val Ser Ile Glu Asp Leu Trp Arg Asn  
 805                      810                      815  
 Glu Gln Phe Trp Val Ile Gly Gly Val Ser Ala His Leu Phe Ala Val  
 820                      825                      830  
 Phe Gln Gly Phe Leu Lys Met Leu Ala Gly Ile Asp Thr Asn Phe Thr  
 835                      840                      845  
 Val Thr Ala Lys Ala Ala Asp Asp Ala Asp Phe Gly Glu Leu Tyr Ile  
 850                      855                      860  
 Val Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Leu Leu Ile Val  
 865                      870                      875                      880  
 Asn Met Val Gly Val Val Ala Gly Phe Ser Asp Ala Leu Asn Lys Gly  
 885                      890                      895  
 Tyr Glu Ala Trp Gly Pro Leu Phe Gly Lys Val Phe Phe Ser Phe Trp  
 900                      905                      910  
 Val Ile Leu His Leu Tyr Pro Phe Leu Lys Gly Leu Met Gly Arg Gln  
 915                      920                      925  
 Asn Arg Thr Pro Thr Ile Val Val Leu Trp Ser Val Leu Leu Ala Ser  
 930                      935                      940  
 Val Phe Ser Leu Val Trp Val Arg Ile Asn Pro Phe Val Ser Thr Ala  
 945                      950                      955                      960  
 Asp Ser Thr Thr Val Ser Gln Ser Cys Ile Ser Ile Asp Cys  
 965                      970

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3311 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA to mRNA  
(v1) ORIGINAL SOURCE:

(A) ORGANISM: *Gossypium hirsutum* L.  
(C) INDIVIDUAL ISOLATE: Coker312

(1x) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 23..3142

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CMTGGTCT TTTGGTTTGG CC ATG GCT TCA ACC ACC ATG GGC GCT GGC TTT 52  
Met Ala Ser Thr Thr Met Ala Ala Gly Phe

[illegible]

TCA ACG AAA ATA TGC AGG GTG TGT GGG GAT AAG ATC GGG CAA AAG GAA 148  
Ser Thr Lys Ile Cys Arg Val Cys Gly Asp Lys Ile Gly Gln Lys Glu  
30 35 40

AAC GGA CAA CCG TTC GTG GCT TGT CAT GTC TGT GCT TTC CCG GTT TGC 196  
Asn Gly Gln Pro Phe Val Ala Cys His Val Cys Ala Phe Pro Val Cys  
45 50 55

CGT OCT TGT TAT GAA TAT GAA AGG AGT GAA GGA AAC CAG TGC TGT OCT 244  
Arg Pro Cys Tyr Glu Tyr Glu Arg Ser Glu Gly Asn Gln Cys Cys Pro  
60 65 70

CAG TGC AAT ACT GGC TAT AAG GGT CAC AAA GGT AGT CCA AGA ATT TCA 292  
Gln Cys Asn Thr Arg Tyr Lys Arg His Lys Gly Ser Pro Arg Ile Ser  
75 80 85 90

GGA GAT GAA GAA GAT GAT TCA GAT CAA GAT GAT TTT GAT GAT GAA TTT 340  
Gly Asp Glu Glu Asp Asp Ser Asp Gln Asp Asp Phe Asp Asp Glu Phe  
95 100 105

CAG ATT AAG AAC CGC AAG GAT GAC TCC CAT CCA CAA CAT GAA AAT GAG 388  
Gln Ile Lys Asn Arg Lys Asp Asp Ser His Pro Gln His Glu Asn Glu  
110 115 120

GAA TAT AAT AAT AAT AAT CAT CAA TGG CAT CCC AAT GGT CAA GCT TTC 436  
Glu Tyr Asn Asn Asn Asn His Gln Trp His Pro Asn Gly Gln Ala Phe  
125 130 135

TCA GTT GGC GGA AGC ACG GCG GGG AAG GAT TTG GAA GGG GAT AAA GAG 484  
Ser Val Ala Gly Ser Thr Ala Gly Lys Asp Leu Glu Gly Asp Lys Glu  
140 145 150

5.5

	365	370	375	
	ATT GAT TAT TTG AAG GAC AAG GTC CAT OCT AGC TTT GTT AAA GAA OGG			1204
5	Ile Asp Tyr Leu Lys Asp Lys Val His Pro Ser Phe Val Lys Glu Arg			
	380	385	390	
	AGA GGC ATG AAA AGG GAA TAT GAA GAA TTT AAA GTA AGG ATC AAT GCA			1252
	Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala			
10	395	400	405	410
	TTA GTA GCA AAA GCT CAG AAG AAA CCA GAA GAA GGA TGG GTG ATG CAA			1300
	Leu Val Ala Lys Ala Gln Lys Lys Pro Glu Glu Gly Trp Val Met Gln			
	415	420	425	
15	GAT GGC ACC CCA TGG CCC GGA AAT AAC ACT OGT GAT CAT OCT GGA ATG			1348
	Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Arg Asp His Pro Gly Met			
	430	435	440	
	ATT CAG GTC TAT CTA GGA AGT GCC GGT GCA CTC GAT GTG GAT GGC AAA			1396
20	Ile Gln Val Tyr Leu Gly Ser Ala Gly Ala Leu Asp Val Asp Gly Lys			
	445	450	455	
	GAG CTG OCT CGA CTT GTC TAT GTT TCT CGT GAG AAA CGA OCT GGT TAT			1444
	Glu Leu Pro Arg Leu Val Tyr Val Ser Arg Glu Lys Arg Pro Gly Tyr			
25	460	465	470	
	CAG CAC CAT AAG AAA GCC GGT GCT GAG AAT GCT CTG GTT CGA GTT TCT			1492
	Gln His His Lys Lys Ala Gly Ala Glu Asn Ala Leu Val Arg Val Ser			
	475	480	485	490
	GCA GTG CTT ACT AAT GCA CCC TTC ATA TTG AAT CTG GAT TGT GAT CAT			1540
30	Ala Val Leu Thr Asn Ala Pro Phe Ile Leu Asn Leu Asp Cys Asp His			
	495	500	505	
	TAC ATC AAC AAT AGC AAG GCC ATG AGG GAA GCG ATG TGC TTT TTA ATG			1588
	Tyr Ile Asn Asn Ser Lys Ala Met Arg Glu Ala Met Cys Phe Leu Met			
35	510	515	520	
	GAT OCT CAG TTT GGA AAG AAG CTT TGT TAT GTT CAA TTT CCA CAG AGA			1636
	Asp Pro Gln Phe Gly Lys Lys Leu Cys Tyr Val Gln Phe Pro Gln Arg			
	525	530	535	
40	TTT GAT GGT ATT GAT OGT CAT GAT CGA TAT GCT AAT CGA AAT GTT GTC			1684
	Phe Asp Gly Ile Asp Arg His Asp Arg Tyr Ala Asn Arg Asn Val Val			
	540	545	550	
	TTC TTT GAT ATC AAC ATG TTG GGA TTA GAT GGA CTT CAA GGC OCT GTA			1732
45	Phe Phe Asp Ile Asn Met Leu Gly Leu Asp Gly Leu Gln Gly Pro Val			
	555	560	565	570
	TAT GTA GGC ACA GGG TGT GTT TTC AAC AGG CAG GCA TTG TAT GGC TAC			1780
	Tyr Val Gly Thr Gly Cys Val Phe Asn Arg Gln Ala Leu Tyr Gly Tyr			
50	575	580	585	
	GAT CCA CCA GTC TCT GAG AAA CGA CCA AAG ATG ACA TGT GAT TGC TGG			1828

55



	Asp	Pro	Pro	Val	Ser	Glu	Lys	Arg	Pro	Lys	Met	Thr	Cys	Asp	Cys	Trp	
				590					595					600			
5	OCT	TCT	TGG	TGT	TGC	TGT	TGT	TGC	GGA	GGT	TCT	AGG	AAG	AAA	TCA	AAG	1876
	Pro	Ser	Trp	Cys	Cys	Cys	Cys	Cys	Gly	Gly	Ser	Arg	Lys	Lys	Ser	Lys	
				605					610					615			
	AAG	AAA	GGT	GAA	AAG	AAG	GGC	TTA	CTC	GGA	GGT	CTT	TTA	TAC	GGA	AAA	1924
10	Lys	Lys	Gly	Glu	Lys	Lys	Gly	Leu	Leu	Gly	Gly	Leu	Leu	Tyr	Gly	Lys	
				620					625					630			
	AAG	AAG	AAG	ATG	ATG	GGC	AAA	AAC	TAT	GTG	AAA	AAA	GGG	TCT	GCA	OCA	1972
	Lys	Lys	Lys	Met	Met	Gly	Lys	Asn	Tyr	Val	Lys	Lys	Gly	Ser	Ala	Pro	
15				635					640					645		650	
	GTC	TTT	GAT	CTC	GAA	GAA	ATC	GAA	GAA	GGG	CTT	GAA	GGA	TAC	GAA	GAA	2020
	Val	Phe	Asp	Leu	Glu	Glu	Ile	Glu	Glu	Gly	Leu	Glu	Gly	Tyr	Glu	Glu	
				655										660		665	
20	TTG	GAG	AAA	TOG	ACA	TTA	ATG	TOG	CAG	AAG	AAT	TTC	GAG	AAA	OGA	TTC	2068
	Leu	Glu	Lys	Ser	Thr	Leu	Met	Ser	Gln	Lys	Asn	Phe	Glu	Lys	Arg	Phe	
				670										675		680	
	GGA	CAA	TCA	COG	GTT	TTC	ATT	GOC	TCA	ACT	TTG	ATG	GAA	AAT	GGT	GGC	2116
25	Gly	Gln	Ser	Pro	Val	Phe	Ile	Ala	Ser	Thr	Leu	Met	Glu	Asn	Gly	Gly	
				685										690		695	
	CTT	OCT	GAA	GGA	ACT	AAT	TOC	ACA	TCA	CTG	ATT	AAA	GAG	GOC	ATT	CAC	2164
	Leu	Pro	Glu	Gly	Thr	Asn	Ser	Thr	Ser	Leu	Ile	Lys	Glu	Ala	Ile	His	
30				700					705					710			
	GTA	ATT	AGC	TGT	GGT	TAT	GAA	GAA	AAA	ACT	GAG	TGG	GGC	AAA	GAG	ATC	2212
	Val	Ile	Ser	Cys	Gly	Tyr	Glu	Glu	Lys	Thr	Glu	Trp	Gly	Lys	Glu	Ile	
				715					720					725		730	
35	GGA	TGG	ATT	TAT	GGG	TOG	GTG	ACG	GAA	GAT	ATA	TTA	ACA	GGT	TTC	AAG	2260
	Gly	Trp	Ile	Tyr	Gly	Ser	Val	Thr	Glu	Asp	Ile	Leu	Thr	Gly	Phe	Lys	
				735										740		745	
	ATG	CAT	TGT	AGA	GGG	TGG	AAA	TOG	GTT	TAT	TGT	GTA	COG	AAA	AGA	COG	2308
40	Met	His	Cys	Arg	Gly	Trp	Lys	Ser	Val	Tyr	Cys	Val	Pro	Lys	Arg	Pro	
				750										755		760	
	GCA	TTC	AAA	GGG	TOC	GCT	OCA	ATC	AAT	CTC	TOG	GAT	OGG	TTG	CAC	CAA	2356
	Ala	Phe	Lys	Gly	Ser	Ala	Pro	Ile	Asn	Leu	Ser	Asp	Arg	Leu	His	Gln	
				765					770					775			
45	GTT	TTG	AGA	TGG	GCA	CTT	GGT	TCT	GTA	GAA	ATT	TTC	CTT	AGT	CGT	CAC	2404
	Val	Leu	Arg	Trp	Ala	Leu	Gly	Ser	Val	Glu	Ile	Phe	Leu	Ser	Arg	His	
				780					785					790			
50	TGT	OCA	CTT	TGG	TAT	GGT	TAT	GGT	GGA	AAA	CTG	AAA	TGG	CTC	GAG	AGG	2452
	Cys	Pro	Leu	Trp	Tyr	Gly	Tyr	Gly	Gly	Lys	Leu	Lys	Trp	Leu	Glu	Arg	
				795				800					805			810	

55

	CTT GCT TAT ATC AAC ACC ATT GTT TAC OCT TTC AOC TOG ATC OCT TTA	2500
	Leu Ala Tyr Ile Asn Thr Ile Val Tyr Pro Phe Thr Ser Ile Pro Leu	
5	815 820 825	
	CTC GOC TAT TGT ACT ATT OCA GCT GTT TGT CTT CTC AOC GGC AAA TTC	2548
	Leu Ala Tyr Cys Thr Ile Pro Ala Val Cys Leu Leu Thr Gly Lys Phe	
	830 835 840	
10	ATC ATT OCA ACT CTA AGC AAC CTT ACA AGT GTG TGG TTC TTG GCA CTT	2596
	Ile Ile Pro Thr Leu Ser Asn Leu Thr Ser Val Trp Phe Leu Ala Leu	
	845 850 855	
	TTC CTC TOC ATC ATT GCA ACT GGA GTG CTT GAA CTT OGA TGG AGC GGG	2644
15	Phe Leu Ser Ile Ile Ala Thr Gly Val Leu Glu Leu Arg Trp Ser Gly	
	860 865 870	
	GTT AGC ATC CAA GAC TGG TGG GGC AAT GAA CAA TTC TGG GTG ATC GGA	2692
	Val Ser Ile Gln Asp Trp Trp Arg Asn Glu Gln Phe Trp Val Ile Gly	
20	875 880 885 890	
	GGT GTC TOC GOC CAT CTT TTT GCT GTC TTC CAG GGC CTC CTC AAA GTC	2740
	Gly Val Ser Ala His Leu Phe Ala Val Phe Gln Gly Leu Leu Lys Val	
	895 900 905	
25	CTA GCT GGA GTA GAC ACC AAC TTC ACC GTA ACA GCA AAA GCA GCA GAC	2788
	Leu Ala Gly Val Asp Thr Asn Phe Thr Val Thr Ala Lys Ala Ala Asp	
	910 915 920	
	GAT ACA GAA TTC GGT GAA CTT TAT CTC TTC AAA TGG ACA ACT CTC TTA	2836
30	Asp Thr Glu Phe Gly Glu Leu Tyr Leu Phe Lys Trp Thr Thr Leu Leu	
	925 930 935	
	ATC OCT OOC ACA ACT CTG ATA ATA CTG AAC ATG GTC GGA GTC GTG GOC	2884
	Ile Pro Pro Thr Thr Leu Ile Ile Leu Asn Met Val Gly Val Val Ala	
	940 945 950	
35	GGA GTT TCA GAC GCA ATC AAC AAC GGC TAT GGT TCA TGG GGT OCA TTG	2932
	Gly Val Ser Asp Ala Ile Asn Asn Gly Tyr Gly Ser Trp Gly Pro Leu	
	955 960 965 970	
40	TTC GGC AAA CTG TTC TTC GCA TTC TGG GTC ATT CTT CAT CTT TAC OCA	2980
	Phe Gly Lys Leu Phe Phe Ala Phe Trp Val Ile Leu His Leu Tyr Pro	
	975 980 985	
	TTC CTC AAA GGT TTG ATG GGG AGA CAA AAC AGG ACG OOC AOC ATT GTT	3028
45	Phe Leu Lys Gly Leu Met Gly Arg Gln Asn Arg Thr Pro Thr Ile Val	
	990 995 1000	
	GTG CTT TGG TOC ATA CTT TTG GCA TOG ATT TTC TCA CTG GTT TGG GTA	3076
	Val Leu Trp Ser Ile Leu Leu Ala Ser Ile Phe Ser Leu Val Trp Val	
	1005 1010 1015	
50	OGG ATC GAT OOC TTC TTG OOC AAA CAA ACA GGT OCA GTT CTT AAA CAA	3124
	Arg Ile Asp Pro Phe Leu Pro Lys Gln Thr Gly Pro Val Leu Lys Gln	

55

1020 1025 1030  
 TGT GGC GTG GAG TGC TAAATGGTGT TTTACAAACC TTCTTTATTA TTTTATTTTC 3179  
 Cys Gly Val Glu Cys  
 1035  
 OCTTTTGGCC ACTACTGTTG ATTTGCTGTG ATTCTAAAAG GGATTTATCT TGTTTGTAAA 3239  
 AAGTCTCTTA TGATTTTGTG GGTCAATT AATTCTATA TGGTAAAAAA ATATTTCTTT 3299  
 AAATTAAC TA 3311

## (2) INFORMATION FOR SEQ ID NO: 4:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1039 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Ala Ser Thr Thr Met Ala Ala Gly Phe Gly Ser Leu Ala Val Asp  
 1 5 10 15  
 Glu Asn Arg Gly Ser Ser Thr His Gln Ser Ser Thr Lys Ile Cys Arg  
 20 25 30  
 Val Cys Gly Asp Lys Ile Gly Gln Lys Glu Asn Gly Gln Pro Phe Val  
 35 40 45  
 Ala Cys His Val Cys Ala Phe Pro Val Cys Arg Pro Cys Tyr Glu Tyr  
 50 55 60  
 Glu Arg Ser Glu Gly Asn Gln Cys Cys Pro Gln Cys Asn Thr Arg Tyr  
 65 70 75 80  
 Lys Arg His Lys Gly Ser Pro Arg Ile Ser Gly Asp Glu Glu Asp Asp  
 85 90 95  
 Ser Asp Gln Asp Asp Phe Asp Asp Glu Phe Gln Ile Lys Asn Arg Lys  
 100 105 110  
 Asp Asp Ser His Pro Gln His Glu Asn Glu Glu Tyr Asn Asn Asn Asn  
 115 120 125  
 His Gln Trp His Pro Asn Gly Gln Ala Phe Ser Val Ala Gly Ser Thr  
 130 135 140  
 Ala Gly Lys Asp Leu Glu Gly Asp Lys Glu Ile Tyr Gly Ser Glu Glu  
 145 150 155 160  
 Trp Lys Glu Arg Val Glu Lys Trp Lys Val Arg Gln Glu Lys Arg Gly  
 165 170 175  
 Leu Val Ser Asn Asp Asn Gly Gly Asn Asp Pro Pro Glu Glu Asp Asp  
 180 185 190  
 Tyr Leu Leu Ala Glu Ala Arg Gln Pro Leu Trp Arg Lys Val Pro Ile  
 195 200 205

Ser Ser Ser Leu Ile Ser Pro Tyr Arg Ile Val Ile Val Leu Arg Phe  
 210 215 220  
 5 Phe Ile Leu Ala Phe Phe Leu Arg Phe Arg Ile Leu Thr Pro Ala Tyr  
 225 230 235 240  
 Asp Ala Tyr Pro Leu Trp Leu Ile Ser Val Ile Cys Glu Val Trp Phe  
 245 250 255  
 10 Ala Phe Ser Trp Ile Leu Asp Gln Phe Pro Lys Trp Phe Pro Ile Thr  
 260 265 270  
 Arg Glu Thr Tyr Leu Asp Arg Leu Ser Leu Arg Phe Glu Arg Glu Gly  
 275 280 285  
 15 Glu Pro Asn Gln Leu Gly Pro Val Asp Val Phe Val Ser Thr Val Asp  
 290 295 300  
 Leu Leu Lys Glu Pro Pro Ile Ile Thr Ala Asn Ala Val Leu Ser Ile  
 305 310 315 320  
 20 Leu Ala Val Asp Tyr Pro Val Glu Lys Val Cys Cys Tyr Val Ser Asp  
 325 330 335  
 Asp Gly Ala Ser Met Leu Leu Phe Asp Ser Leu Ser Glu Thr Ala Glu  
 340 345 350  
 25 Phe Ala Arg Arg Trp Val Pro Phe Cys Lys Lys His Asn Val Glu Pro  
 355 360 365  
 Arg Ala Pro Glu Phe Tyr Phe Asn Glu Lys Ile Asp Tyr Leu Lys Asp  
 370 375 380  
 30 Lys Val His Pro Ser Phe Val Lys Glu Arg Arg Ala Met Lys Arg Glu  
 385 390 395 400  
 Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala Gln  
 405 410 415  
 35 Lys Lys Pro Glu Glu Gly Trp Val Met Gln Asp Gly Thr Pro Trp Pro  
 420 425 430  
 Gly Asn Asn Thr Arg Asp His Pro Gly Met Ile Gln Val Tyr Leu Gly  
 435 440 445  
 40 Ser Ala Gly Ala Leu Asp Val Asp Gly Lys Glu Leu Pro Arg Leu Val  
 450 455 460  
 Tyr Val Ser Arg Glu Lys Arg Pro Gly Tyr Gln His His Lys Lys Ala  
 465 470 475 480  
 45 Gly Ala Glu Asn Ala Leu Val Arg Val Ser Ala Val Leu Thr Asn Ala  
 485 490 495  
 Pro Phe Ile Leu Asn Leu Asp Cys Asp His Tyr Ile Asn Asn Ser Lys  
 500 505 510  
 50 Ala Met Arg Glu Ala Met Cys Phe Leu Met Asp Pro Gln Phe Gly Lys  
 515 520 525  
 55 Lys Leu Cys Tyr Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp Arg

530 535 540  
 His Asp Arg Tyr Ala Asn Arg Asn Val Val Phe Phe Asp Ile Asn Met  
 545 550 555 560  
 Leu Gly Leu Asp Gly Leu Gln Gly Pro Val Tyr Val Gly Thr Gly Cys  
 565 570 575  
 Val Phe Asn Arg Gln Ala Leu Tyr Gly Tyr Asp Pro Pro Val Ser Glu  
 580 585 590  
 Lys Arg Pro Lys Met Thr Cys Asp Cys Trp Pro Ser Trp Cys Cys Cys  
 595 600 605  
 Cys Cys Gly Gly Ser Arg Lys Lys Ser Lys Lys Lys Gly Glu Lys Lys  
 610 615 620  
 Gly Leu Leu Gly Gly Leu Leu Tyr Gly Lys Lys Lys Lys Met Met Gly  
 625 630 635 640  
 Lys Asn Tyr Val Lys Lys Gly Ser Ala Pro Val Phe Asp Leu Glu Glu  
 645 650 655  
 Ile Glu Glu Gly Leu Glu Gly Tyr Glu Glu Leu Glu Lys Ser Thr Leu  
 660 665 670  
 Met Ser Gln Lys Asn Phe Glu Lys Arg Phe Gly Gln Ser Pro Val Phe  
 675 680 685  
 Ile Ala Ser Thr Leu Met Glu Asn Gly Gly Leu Pro Glu Gly Thr Asn  
 690 695 700  
 Ser Thr Ser Leu Ile Lys Glu Ala Ile His Val Ile Ser Cys Gly Tyr  
 705 710 715 720  
 Glu Glu Lys Thr Glu Trp Gly Lys Glu Ile Gly Trp Ile Tyr Gly Ser  
 725 730 735  
 Val Thr Glu Asp Ile Leu Thr Gly Phe Lys Met His Cys Arg Gly Trp  
 740 745 750  
 Lys Ser Val Tyr Cys Val Pro Lys Arg Pro Ala Phe Lys Gly Ser Ala  
 755 760 765  
 Pro Ile Asn Leu Ser Asp Arg Leu His Gln Val Leu Arg Trp Ala Leu  
 770 775 780  
 Gly Ser Val Glu Ile Phe Leu Ser Arg His Cys Pro Leu Trp Tyr Gly  
 785 790 795 800  
 Tyr Gly Gly Lys Leu Lys Trp Leu Glu Arg Leu Ala Tyr Ile Asn Thr  
 805 810 815  
 Ile Val Tyr Pro Phe Thr Ser Ile Pro Leu Leu Ala Tyr Cys Thr Ile  
 820 825 830  
 Pro Ala Val Cys Leu Leu Thr Gly Lys Phe Ile Ile Pro Thr Leu Ser  
 835 840 845  
 Asn Leu Thr Ser Val Trp Phe Leu Ala Leu Phe Leu Ser Ile Ile Ala  
 850 855 860

Thr Gly Val Leu Glu Leu Arg Trp Ser Gly Val Ser Ile Gln Asp Trp  
 865 870 875 880  
 Trp Arg Asn Glu Gln Phe Trp Val Ile Gly Gly Val Ser Ala His Leu  
 885 890 895  
 Phe Ala Val Phe Gln Gly Leu Leu Lys Val Leu Ala Gly Val Asp Thr  
 900 905 910  
 Asn Phe Thr Val Thr Ala Lys Ala Ala Asp Asp Thr Glu Phe Gly Glu  
 915 920 925  
 Leu Tyr Leu Phe Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Leu  
 930 935 940  
 Ile Ile Leu Asn Met Val Gly Val Val Ala Gly Val Ser Asp Ala Ile  
 945 950 955 960  
 Asn Asn Gly Tyr Gly Ser Trp Gly Pro Leu Phe Gly Lys Leu Phe Phe  
 965 970 975  
 Ala Phe Trp Val Ile Leu His Leu Tyr Pro Phe Leu Lys Gly Leu Met  
 980 985 990  
 Gly Arg Gln Asn Arg Thr Pro Thr Ile Val Val Leu Trp Ser Ile Leu  
 995 1000 1005  
 Leu Ala Ser Ile Phe Ser Leu Val Trp Val Arg Ile Asp Pro Phe Leu  
 1010 1015 1020  
 Pro Lys Gln Thr Gly Pro Val Leu Lys Gln Cys Gly Val Glu Cys  
 1025 1030 1035

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2033 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA to mRNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Gossypium hirsutum* L.
- (C) INDIVIDUAL ISOLATE: Coker312

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1857

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCG ACA TTC GTG AAG GAG CGT CGA GCT ATG AAG AGA GAA TAT GAA GAA 48  
 Pro Thr Phe Val Lys Glu Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu  
 1 5 10 15  
 TTC AAG GTT AGG ATA AAT GCA CTT GTA GGC AAA GGC CAA AAG GTT CCT 96

	Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala Gln Lys Val Pro	
	20 25 30	
5	CCA GAA GGG TGG ATC ATG CAA GAT GGG ACA CCA TGG OCA GGA AAC AAT	144
	Pro Glu Gly Trp Ile Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn	
	35 40 45	
10	ACT AAA GAT CAC CCT GGT ATG ATT CAA GTA TTT CTC GGT CAA AGT GGA	192
	Thr Lys Asp His Pro Gly Met Ile Gln Val Phe Leu Gly Gln Ser Gly	
	50 55 60	
	GGC CAT GAT AOC GAA GGA AAT GAG CTT OCT CGT CTC GTC TAT GTA TCT	240
	Gly His Asp Thr Glu Gly Asn Glu Leu Pro Arg Leu Val Tyr Val Ser	
15	65 70 75 80	
	CGA GAG AAA AGG OCT GGT TTC TTG CAT CAC AAG AAA GCT GGT GOC ATG	288
	Arg Glu Lys Arg Pro Gly Phe Leu His His Lys Lys Ala Gly Ala Met	
	85 90 95	
20	AAC GOC CTT GTT CCG GTC TOG GGG GTG CTC ACA AAT GCT OCT TTT ATG	336
	Asn Ala Leu Val Arg Val Ser Gly Val Leu Thr Asn Ala Pro Phe Met	
	100 105 110	
25	TTG AAC TTG GAT TGT GAC CAT TAT TTA AAT AAC AGC AAG GCT GTA AGA	384
	Leu Asn Leu Asp Cys Asp His Tyr Leu Asn Asn Ser Lys Ala Val Arg	
	115 120 125	
	GAG GCT ATG TGT TTC TTG ATG GAC OCT CAA ATT GGA AGA AAG GTT TGC	432
	Glu Ala Met Cys Phe Leu Met Asp Pro Gln Ile Gly Arg Lys Val Cys	
30	130 135 140	
	TAT GTC CAA TTC OCT CAA CGT TTC GAT GGT ATT GAT AGA CAT GAT OGA	480
	Tyr Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp Arg His Asp Arg	
	145 150 155 160	
35	TAT GOC AAT CCG AAC ACA GTT TTC TTT GAT ATT AAC ATG AAA GGT CTA	528
	Tyr Ala Asn Arg Asn Thr Val Phe Phe Asp Ile Asn Met Lys Gly Leu	
	165 170 175	
40	GAT GGT ATA CAA GGC OCT GTA TAT GTC GGC ACG GGG TGT GTT TTC AGA	576
	Asp Gly Ile Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Val Phe Arg	
	180 185 190	
	AGG CAA GCT CTT TAT GGT TAT GAA OCT OCA AAG GGA OCT AAG GGC CCG	624
	Arg Gln Ala Leu Tyr Gly Tyr Glu Pro Pro Lys Gly Pro Lys Arg Pro	
45	195 200 205	
	AAA ATG GTA AOC TGT GGT TGC TGC OCT TGT TTT GGA GGC GGC AGA AAG	672
	Lys Met Val Thr Cys Gly Cys Cys Pro Cys Phe Gly Arg Arg Arg Lys	
	210 215 220	
50	GAC AAA AAG CAC TCT AAG GAT GGT GGA AAT GCA AAT GGT CTA AGC CTA	720
	Asp Lys Lys His Ser Lys Asp Gly Gly Asn Ala Asn Gly Leu Ser Leu	
	225 230 235 240	

55

	GAA GCA GCC AAA GAT GAC AAG GAG TTA TTG ATG TOC CAC ATG AAC TTT	768
	Glu Ala Ala Lys Asp Asp Lys Glu Leu Leu Met Ser His Met Asn Phe	
5	245 250 255	
	GAA AAG AAA TTT GGA CAA TCA GOC ATT TTT GTA ACT TCA ACA CTG ATG	816
	Glu Lys Lys Phe Gly Gln Ser Ala Ile Phe Val Thr Ser Thr Leu Met	
	260 265 270	
10	GAA CAA GGT GGT GTC OCT OCT TCT TCA AGC OOC GCA GCT TTG CTC AAA	864
	Glu Gln Gly Gly Val Pro Pro Ser Ser Pro Ala Ala Leu Leu Lys	
	275 280 285	
	GAA GOC ATT CAT GTA ATT AGT TGT GGT TAT GAA GAC AAA ACA GAA TGG	912
	Glu Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp	
15	290 295 300	
	GGA AGC GAG CTT GGC TGG ATT TAC GGC TGG ATT ACA GAA GAT ATC TTA	960
	Gly Ser Glu Leu Gly Trp Ile Tyr Gly Ser Ile Thr Glu Asp Ile Leu	
	305 310 315 320	
20	ACA GGA TTC AAG ATG CAT TGC OGT GGA TGG AGA TCA ATA TAC TGC ATG	1008
	Thr Gly Phe Lys Met His Cys Arg Gly Trp Arg Ser Ile Tyr Cys Met	
	325 330 335	
	CCA AAG TTG OCT GCA TTC AAG GGT TCA GCT OOC ATC AAT CTA TGG GAT	1056
25	Pro Lys Leu Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp	
	340 345 350	
	OGT CTA AAC CAA GTC CTT CGA TGG GCA CTC GGT TCT GTT GAA ATT TTC	1104
	Arg Leu Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu Ile Phe	
30	355 360 365	
	TTT AGT CAT CAT TGC CCA GCA TGG TAT GGT TTC AAG GGA GGA AAG CTA	1152
	Phe Ser His His Cys Pro Ala Trp Tyr Gly Phe Lys Gly Gly Lys Leu	
	370 375 380	
35	AAA TGG CTT GAA CGA TTC GCA TAT GTC AAC ACA ACC ATC TAC OOC TTC	1200
	Lys Trp Leu Glu Arg Phe Ala Tyr Val Asn Thr Thr Ile Tyr Pro Phe	
	385 390 395 400	
	ACA TCT TTA CCA CTT CTC GOC TAT TGT ACC CTA OCG GCA ATC TGT TTA	1248
40	Thr Ser Leu Pro Leu Leu Ala Tyr Cys Thr Leu Pro Ala Ile Cys Leu	
	405 410 415	
	CTT ACC GAT AAA TTT ATC ATG CCA CCG ATA AGC ACC TTT GCA AGT CTA	1296
	Leu Thr Asp Lys Phe Ile Met Pro Pro Ile Ser Thr Phe Ala Ser Leu	
	420 425 430	
45	TTC TTC ATT GOC TTG TTT CTT TCA ATC TTT GCA ACT GGT ATT CTC GAG	1344
	Phe Phe Ile Ala Leu Phe Leu Ser Ile Phe Ala Thr Gly Ile Leu Glu	
	435 440 445	
50	CTA AGG TGG AGT GGA GTA AGC ATT GAA GAA TGG TGG AGG AAT GAG CAA	1392
	Leu Arg Trp Ser Gly Val Ser Ile Glu Glu Trp Trp Arg Asn Glu Gln	

55



	450	455	460	
	TTT TGG GTC ATC GGT GGC ATT TCG GCA CAT TTG TTC GCT GTT ATC CAA			1440
5	Phe Trp Val Ile Gly Gly Ile Ser Ala His Leu Phe Ala Val Ile Gln			
	465	470	475	480
	GGC TTG TTG AAA GTT CTA GCT GGT ATT GAC ACT AAT TTC ACT GTC ACA			1488
	Gly Leu Leu Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr			
10	485	490	495	
	TOC AAG GCA ACT GAT GAC GAG GAG TTC GGG GAA TTG TAT ACT TTC AAA			1536
	Ser Lys Ala Thr Asp Asp Glu Glu Phe Gly Glu Leu Tyr Thr Phe Lys			
	500	505	510	
15	TGG ACA ACC CTT CTA ATT OCT OCT ACT ACC GTC TTA ATC ATC AAT TTA			1584
	Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Val Leu Ile Ile Asn Leu			
	515	520	525	
	GTC GGT GTC GTT GCA GGC ATC TCG GAT GGC ATA AAC AAT GGA TAC CAA			1632
20	Val Gly Val Val Ala Gly Ile Ser Asp Ala Ile Asn Asn Gly Tyr Gln			
	530	535	540	
	TCA TGG GGA OCT CTT TTT GGG AAG CTC TTC TTC TCT TTC TGG GTG ATT			1680
	Ser Trp Gly Pro Leu Phe Gly Lys Leu Phe Phe Ser Phe Trp Val Ile			
25	545	550	555	560
	GTC CAT CTC TAT CCA TTC CTC AAA GGT TTA ATG GGG AGA CAA AAC CGG			1728
	Val His Leu Tyr Pro Phe Leu Lys Gly Leu Met Gly Arg Gln Asn Arg			
	565	570	575	
30	ACA CCA ACC ATT GTT GTT ATA TGG TCA GTG CTA TTG GCT TCA ATC TTC			1776
	Thr Pro Thr Ile Val Val Ile Trp Ser Val Leu Leu Ala Ser Ile Phe			
	580	585	590	
	TOC TTG CTT TGG GTC CGA ATT GAT CCA TTT GTG ATG AAA AOC AAA GGA			1824
35	Ser Leu Leu Trp Val Arg Ile Asp Pro Phe Val Met Lys Thr Lys Gly			
	595	600	605	
	OCA GAC ACT ACA ATG TGT GGC ATT AAC TGT TGAAAAAAAA TCATCTTGCG			1874
	Pro Asp Thr Thr Met Cys Gly Ile Asn Cys			
	610	615		
40	TGGTTCTTTT AGATTATGGT ATGTGATGTA TGAACAAACA AGAATGGAGA TGCACAAGAC			1934
	AGAATAAAAT TAGAGTGAAA GTTTTGTGTA GTTATATATT CATTCTACCA ACTATAAGTT			1994
	TTGTCATTCA ATTGAAAATA GCTCAACTTT GTGATCAAA			2033

## (2) INFORMATION FOR SEQ ID NO: 6:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 618 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: C-terminal fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

5      Pro Thr Phe Val Lys Glu Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu  
       1                                5                                10                                15  
     Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala Gln Lys Val Pro  
                               20                                25                                30  
 10     Pro Glu Gly Trp Ile Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn  
                               35                                40                                45  
     Thr Lys Asp His Pro Gly Met Ile Gln Val Phe Leu Gly Gln Ser Gly  
                               50                                55                                60  
 15     Gly His Asp Thr Glu Gly Asn Glu Leu Pro Arg Leu Val Tyr Val Ser  
                               65                                70                                75                                80  
     Arg Glu Lys Arg Pro Gly Phe Leu His His Lys Lys Ala Gly Ala Met  
                               85                                90                                95  
 20     Asn Ala Leu Val Arg Val Ser Gly Val Leu Thr Asn Ala Pro Phe Met  
                               100                                105                                110  
     Leu Asn Leu Asp Cys Asp His Tyr Leu Asn Asn Ser Lys Ala Val Arg  
                               115                                120                                125  
 25     Glu Ala Met Cys Phe Leu Met Asp Pro Gln Ile Gly Arg Lys Val Cys  
                               130                                135                                140  
     Tyr Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp Arg His Asp Arg  
                               145                                150                                155                                160  
 30     Tyr Ala Asn Arg Asn Thr Val Phe Phe Asp Ile Asn Met Lys Gly Leu  
                               165                                170                                175  
     Asp Gly Ile Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Val Phe Arg  
                               180                                185                                190  
 35     Arg Gln Ala Leu Tyr Gly Tyr Glu Pro Pro Lys Gly Pro Lys Arg Pro  
                               195                                200                                205  
     Lys Met Val Thr Cys Gly Cys Cys Pro Cys Phe Gly Arg Arg Arg Lys  
                               210                                215                                220  
 40     Asp Lys Lys His Ser Lys Asp Gly Gly Asn Ala Asn Gly Leu Ser Leu  
                               225                                230                                235                                240  
     Glu Ala Ala Lys Asp Asp Lys Glu Leu Leu Met Ser His Met Asn Phe  
                               245                                250                                255  
 45     Glu Lys Lys Phe Gly Gln Ser Ala Ile Phe Val Thr Ser Thr Leu Met  
                               260                                265                                270  
 50     Glu Gln Gly Gly Val Pro Pro Ser Ser Ser Pro Ala Ala Leu Leu Lys  
                               275                                280                                285  
     Glu Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp  
                               290                                295                                300  
 55     Gly Ser Glu Leu Gly Trp Ile Tyr Gly Ser Ile Thr Glu Asp Ile Leu

305                      310                      315                      320  
 Thr Gly Phe Lys Met His Cys Arg Gly Trp Arg Ser Ile Tyr Cys Met  
                                  325                      330                      335  
 Pro Lys Leu Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp  
                                  340                      345                      350  
 Arg Leu Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu Ile Phe  
                                  355                      360                      365  
 Phe Ser His His Cys Pro Ala Trp Tyr Gly Phe Lys Gly Gly Lys Leu  
                                  370                      375                      380  
 Lys Trp Leu Glu Arg Phe Ala Tyr Val Asn Thr Thr Ile Tyr Pro Phe  
 385                      390                      395                      400  
 Thr Ser Leu Pro Leu Leu Ala Tyr Cys Thr Leu Pro Ala Ile Cys Leu  
                                  405                      410                      415  
 Leu Thr Asp Lys Phe Ile Met Pro Pro Ile Ser Thr Phe Ala Ser Leu  
                                  420                      425                      430  
 Phe Phe Ile Ala Leu Phe Leu Ser Ile Phe Ala Thr Gly Ile Leu Glu  
                                  435                      440                      445  
 Leu Arg Trp Ser Gly Val Ser Ile Glu Glu Trp Trp Arg Asn Glu Gln  
 450                      455                      460  
 Phe Trp Val Ile Gly Gly Ile Ser Ala His Leu Phe Ala Val Ile Gln  
 465                      470                      475                      480  
 Gly Leu Leu Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr  
                                  485                      490                      495  
 Ser Lys Ala Thr Asp Asp Glu Glu Phe Gly Glu Leu Tyr Thr Phe Lys  
                                  500                      505                      510  
 Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Val Leu Ile Ile Asn Leu  
                                  515                      520                      525  
 Val Gly Val Val Ala Gly Ile Ser Asp Ala Ile Asn Asn Gly Tyr Gln  
                                  530                      535                      540  
 Ser Trp Gly Pro Leu Phe Gly Lys Leu Phe Phe Ser Phe Trp Val Ile  
 545                      550                      555                      560  
 Val His Leu Tyr Pro Phe Leu Lys Gly Leu Met Gly Arg Gln Asn Arg  
                                  565                      570                      575  
 Thr Pro Thr Ile Val Val Ile Trp Ser Val Leu Leu Ala Ser Ile Phe  
                                  580                      585                      590  
 Ser Leu Leu Trp Val Arg Ile Asp Pro Phe Val Met Lys Thr Lys Gly  
                                  595                      600                      605  
 Pro Asp Thr Thr Met Cys Gly Ile Asn Cys  
                                  610                      615

(2) INFORMATION FOR SEQ ID NO: 7:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1086 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Gossypium hirsutum* L.

(C) INDIVIDUAL ISOLATE: Coker312

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 24..1086

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```

GGCAOGAGCT TTCATATOCCT OCA ATG GAA GGC AGC GGC GGA CTC GTT GOG      50
Met Glu Ala Ser Ala Gly Leu Val Ala
      1              5
GGC TCT CAC AAC CGC AAT GAA CTT GTT GTC ATT CAT GGC CAT GAA GAG      98
Gly Ser His Asn Arg Asn Glu Leu Val Val Ile His Gly His Glu Glu
      10              15              20              25
OCT AAA OCT CTG AAG AAC TTG GAT GGT CAA GTT TGT GAG ATT TGT GGT      146
Pro Lys Pro Leu Lys Asn Leu Asp Gly Gln Val Cys Glu Ile Cys Gly
      30              35              40
GAT GAA ATT GGG TTG ACG GTC GAT GGA GAT CTT TTC GTG GCC TGC AAC      194
Asp Glu Ile Gly Leu Thr Val Asp Gly Asp Leu Phe Val Ala Cys Asn
      45              50              55
GAG TGT GGT TTT CCA GTT TGT AGG OCT TGT TAT GAG TAT GAA AGG AGA      242
Glu Cys Gly Phe Pro Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Arg
      60              65              70
GAA GGG AGT CAA CAA TGT OCT CAA TGC AAA ACT AGA TAC AAG CGT CTC      290
Glu Gly Ser Gln Gln Cys Pro Gln Cys Lys Thr Arg Tyr Lys Arg Leu
      75              80              85
AAG GGG AGT CCG AGG GTG GAG GGA GAT GAA GAT GAA GAG GAT GTG GAT      338
Lys Gly Ser Pro Arg Val Glu Gly Asp Glu Asp Glu Glu Asp Val Asp
      90              95              100              105
GAT ATC GAA CAT GAA TTC AAC ATT GAT GAT GAA CAA AAC AAG TAT AGA      386
Asp Ile Glu His Glu Phe Asn Ile Asp Asp Glu Gln Asn Lys Tyr Arg
      110              115              120
AAT ATC GCT GAA TCG ATG CTT CAT GGA AAG ATG AGC TAC GGG AGA GGC      434
Asn Ile Ala Glu Ser Met Leu His Gly Lys Met Ser Tyr Gly Arg Gly
      125              130              135
OCT GAA GAC GAT GAA GGT TTG CAA ATC CCA CCC GGT TTA GCT GGT GTT      482

```

	Pro Glu Asp Asp Glu Gly Leu Gln Ile Pro Pro Gly Leu Ala Gly Val	
	140 145 150	
5	CGA TCT CGG CCG GTG AGC GGG GAG TTC CCA ATA GGA AGC TCT CTT GCT	530
	Arg Ser Arg Pro Val Ser Gly Glu Phe Pro Ile Gly Ser Ser Leu Ala	
	155 160 165	
	TAT GGG GAA CAC ATG TCA AAT AAA CGA GTT CAT CCA TAT CCT ATG TCT	578
10	Tyr Gly Glu His Met Ser Asn Lys Arg Val His Pro Tyr Pro Met Ser	
	170 175 180 185	
	GAA OCT GGA AGT GCA AGA TGG GAT GAA AAG AAA GAG GGA GGA TGG AGA	626
	Glu Pro Gly Ser Ala Arg Trp Asp Glu Lys Lys Glu Gly Gly Trp Arg	
	190 195 200	
15	GAA AGG ATG GAT GAT TGG AAA ATG CAG CAA GGG AAT TTG GGT OCT GAA	674
	Glu Arg Met Asp Asp Trp Lys Met Gln Gln Gly Asn Leu Gly Pro Glu	
	205 210 215	
20	OCT GAT GAT GOC TAT GAT GCT GAC ATG GCT ATG CTT GAT GAA GCT AGG	722
	Pro Asp Asp Ala Tyr Asp Ala Asp Met Ala Met Leu Asp Glu Ala Arg	
	220 225 230	
	CAG CCA TTG TCA AGG AAA GTG CCA ATT GCA TCG AGC AAA ATC AAT OCT	770
25	Gln Pro Leu Ser Arg Lys Val Pro Ile Ala Ser Ser Lys Ile Asn Pro	
	235 240 245	
	TAT CGT ATG GTG ATT GTG GCT CGT CTA GTT ATC CTT GCT TTC TTT CTT	818
	Tyr Arg Met Val Ile Val Ala Arg Leu Val Ile Leu Ala Phe Phe Leu	
	250 255 260 265	
30	CGC TAT CGG ATT TTG AAC CCG GTA CAT GAT GCA ATT GGG CTT TGG CTA	866
	Arg Tyr Arg Ile Leu Asn Pro Val His Asp Ala Ile Gly Leu Trp Leu	
	270 275 280	
	ACT TCT GTG ATC TGT GAA ATC TGG TTT GOC TTT TCA TGG ATC CTT GAT	914
35	Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Phe Ser Trp Ile Leu Asp	
	285 290 295	
	CAG TTC OCT AAA TGG TTC OCT ATT GAC CGC GAG ACG TAT CTC GAT CGC	962
40	Gln Phe Pro Lys Trp Phe Pro Ile Asp Arg Glu Thr Tyr Leu Asp Arg	
	300 305 310	
	CTT TOC CTC AGG TAT GAG AGG GAA GGT GAG CCC AAC ATG CTT GCT TCT	1010
	Leu Ser Leu Arg Tyr Glu Arg Glu Gly Glu Pro Asn Met Leu Ala Ser	
	315 320 325	
45	GTT GAT ATT TTT GTC AGT ACA GTG GAT CCA TTG AAG GGA CCT OCT CTA	1058
	Val Asp Ile Phe Val Ser Thr Val Asp Pro Leu Lys Gly Pro Pro Leu	
	330 335 340 345	
	GTA ACA GCG AAT ACA GTT CTA TCG ATC T	1086
50	Val Thr Ala Asn Thr Val Leu Ser Ile	
	350	
55		

## (2) INFORMATION FOR SEQ ID NO: 8:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 354 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminal fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Glu Ala Ser Ala Gly Leu Val Ala Gly Ser His Asn Arg Asn Glu  
 1 5 10 15  
 Leu Val Val Ile His Gly His Glu Glu Pro Lys Pro Leu Lys Asn Leu  
 20 25 30  
 Asp Gly Gln Val Cys Glu Ile Cys Gly Asp Glu Ile Gly Leu Thr Val  
 35 40 45  
 Asp Gly Asp Leu Phe Val Ala Cys Asn Glu Cys Gly Phe Pro Val Cys  
 50 55 60  
 Arg Pro Cys Tyr Glu Tyr Glu Arg Arg Glu Gly Ser Gln Gln Cys Pro  
 65 70 75 80  
 Gln Cys Lys Thr Arg Tyr Lys Arg Leu Lys Gly Ser Pro Arg Val Glu  
 85 90 95  
 Gly Asp Glu Asp Glu Glu Asp Val Asp Asp Ile Glu His Glu Phe Asn  
 100 105 110  
 Ile Asp Asp Glu Gln Asn Lys Tyr Arg Asn Ile Ala Glu Ser Met Leu  
 115 120 125  
 His Gly Lys Met Ser Tyr Gly Arg Gly Pro Glu Asp Asp Glu Gly Leu  
 130 135 140  
 Gln Ile Pro Pro Gly Leu Ala Gly Val Arg Ser Arg Pro Val Ser Gly  
 145 150 155 160  
 Glu Phe Pro Ile Gly Ser Ser Leu Ala Tyr Gly Glu His Met Ser Asn  
 165 170 175  
 Lys Arg Val His Pro Tyr Pro Met Ser Glu Pro Gly Ser Ala Arg Trp  
 180 185 190  
 Asp Glu Lys Lys Glu Gly Gly Trp Arg Glu Arg Met Asp Asp Trp Lys  
 195 200 205  
 Met Gln Gln Gly Asn Leu Gly Pro Glu Pro Asp Asp Ala Tyr Asp Ala  
 210 215 220  
 Asp Met Ala Met Leu Asp Glu Ala Arg Gln Pro Leu Ser Arg Lys Val  
 225 230 235 240  
 Pro Ile Ala Ser Ser Lys Ile Asn Pro Tyr Arg Met Val Ile Val Ala  
 245 250 255

Arg Leu Val Ile Leu Ala Phe Phe Leu Arg Tyr Arg Ile Leu Asn Pro  
 260 265 270  
 Val His Asp Ala Ile Gly Leu Trp Leu Thr Ser Val Ile Cys Glu Ile  
 275 280 285  
 Trp Phe Ala Phe Ser Trp Ile Leu Asp Gln Phe Pro Lys Trp Phe Pro  
 290 295 300  
 Ile Asp Arg Glu Thr Tyr Leu Asp Arg Leu Ser Leu Arg Tyr Glu Arg  
 305 310 315 320  
 Glu Gly Glu Pro Asn Met Leu Ala Ser Val Asp Ile Phe Val Ser Thr  
 325 330 335  
 Val Asp Pro Leu Lys Gly Pro Pro Leu Val Thr Ala Asn Thr Val Leu  
 340 345 350  
 Ser Ile

## (2) INFORMATION FOR SEQ ID NO: 9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1000 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA to mRNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Gossypium hirsutum* L.
- (C) INDIVIDUAL ISOLATE: Coker312

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1000

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAC AAA GTC CCG CCG ACA TTC GTG AAG GAG CGT CGA GCT ATG AAG AGA	48
Asp Lys Val Arg Pro Thr Phe Val Lys Glu Arg Arg Ala Met Lys Arg	
1 5 10 15	
GAA TAT GAA GAA TTC AAG GTT AGG ATA AAT GCA CTT GTA GGC AAA GGC	96
Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala	
20 25 30	
CAA AAG GTT OCT CCA GAA GGG TGG ATC ATG CAA GAT GGG ACA CCA TGG	144
Gln Lys Val Pro Pro Glu Gly Trp Ile Met Gln Asp Gly Thr Pro Trp	
35 40 45	
OCA GGA AAC AAT ACT AAA GAT CAC OCT GGT ATG ATT CAA GTA TTT CTC	192
Pro Gly Asn Asn Thr Lys Asp His Pro Gly Met Ile Gln Val Phe Leu	
50 55 60	

BNSDOCID: &lt;EP 0876575A2 L&gt;



275                      280                      285  
 GCT TTG CTC AAA GAA GCC ATT CAT GTA ATT AGT TGT GGT TAT GAA GAC                      912  
 Ala Leu Leu Lys Glu Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp  
 290                      295                      300  
 AAA ACC GAA TGG GGA AGC GAG CTT GGC TGG ATT TAC GGC TCG ATT ACA                      960  
 Lys Thr Glu Trp Gly Ser Glu Leu Gly Trp Ile Tyr Gly Ser Ile Thr  
 305                      310                      315                      320  
 GAA GAT ATC TTA ACA GGT TTC AAG ATG CAT TGC CGT GGA T                      1000  
 Glu Asp Ile Leu Thr Gly Phe Lys Met His Cys Arg Gly  
 325                      330

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 333 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

25 Asp Lys Val Arg Pro Thr Phe Val Lys Glu Arg Arg Ala Met Lys Arg  
 1                      5                      10                      15  
 Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala  
 20                      25                      30  
 30 Gln Lys Val Pro Pro Glu Gly Trp Ile Met Gln Asp Gly Thr Pro Trp  
 35                      40                      45  
 Pro Gly Asn Asn Thr Lys Asp His Pro Gly Met Ile Gln Val Phe Leu  
 50                      55                      60  
 35 Gly Gln Ser Gly Gly His Asp Thr Glu Gly Asn Glu Leu Pro Arg Leu  
 65                      70                      75                      80  
 Val Tyr Val Ser Arg Glu Lys Arg Pro Gly Phe Leu His His Lys Lys  
 85                      90                      95  
 40 Ala Gly Ala Met Asn Ala Leu Val Arg Val Ser Gly Val Leu Thr Asn  
 100                      105                      110  
 Ala Pro Phe Met Leu Asn Leu Asp Cys Asp His Tyr Leu Asn Asn Ser  
 115                      120                      125  
 45 Lys Ala Val Arg Glu Ala Met Cys Phe Leu Met Asp Pro Gln Ile Gly  
 130                      135                      140  
 Arg Lys Val Cys Tyr Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp  
 145                      150                      155                      160  
 50 Arg His Asp Arg Tyr Ala Asn Arg Asn Thr Val Phe Phe Asp Ile Asn  
 165                      170                      175

Met Lys Gly Leu Asp Gly Ile Gln Gly Pro Val Tyr Val Gly Thr Gly  
 180 185 190  
 Cys Val Phe Arg Arg Gln Ala Leu Tyr Gly Tyr Glu Pro Pro Lys Gly  
 195 200 205  
 Pro Lys Arg Pro Lys Met Val Thr Cys Gly Cys Cys Pro Cys Phe Gly  
 210 215 220  
 Arg Arg Arg Lys Asp Lys Lys His Ser Lys Asp Gly Gly Asn Ala Asn  
 225 230 235 240  
 Gly Leu Ser Leu Glu Ala Ala Glu Asp Asp Lys Glu Leu Leu Met Ser  
 245 250 255  
 His Met Asn Phe Glu Lys Lys Phe Gly Gln Ser Ala Ile Phe Val Thr  
 260 265 270  
 Ser Thr Leu Met Glu Gln Gly Gly Val Pro Pro Ser Ser Ser Pro Ala  
 275 280 285  
 Ala Leu Leu Lys Glu Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp  
 290 295 300  
 Lys Thr Glu Trp Gly Ser Glu Leu Gly Trp Ile Tyr Gly Ser Ile Thr  
 305 310 315 320  
 Glu Asp Ile Leu Thr Gly Phe Lys Met His Cys Arg Gly  
 325 330

## (2) INFORMATION FOR SEQ ID NO: 11:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 622 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: C-terminal fragment

## (1x) FEATURE:

(A) NAME/KEY:

(B) LOCATION:

(D) OTHER INFORMATION: Xaa indicates Glu or Lys

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Asp Lys Val Arg Pro Thr Phe Val Lys Glu Arg Arg Ala Met Lys Arg  
 1 5 10 15  
 Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala  
 20 25 30  
 Gln Lys Val Pro Pro Glu Gly Trp Ile Met Gln Asp Gly Thr Pro Trp  
 35 40 45  
 Pro Gly Asn Asn Thr Lys Asp His Pro Gly Met Ile Gln Val Phe Leu  
 50 55 60

Gly Gln Ser Gly Gly His Asp Thr Glu Gly Asn Glu Leu Pro Arg Leu  
 65 70 75 80  
 5 Val Tyr Val Ser Arg Glu Lys Arg Pro Gly Phe Leu His His Lys Lys  
 85 90 95  
 Ala Gly Ala Met Asn Ala Leu Val Arg Val Ser Gly Val Leu Thr Asn  
 100 105 110  
 10 Ala Pro Phe Met Leu Asn Leu Asp Cys Asp His Tyr Leu Asn Asn Ser  
 115 120 125  
 Lys Ala Val Arg Glu Ala Met Cys Phe Leu Met Asp Pro Gln Ile Gly  
 130 135 140  
 15 Arg Lys Val Cys Tyr Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp  
 145 150 155 160  
 Arg His Asp Arg Tyr Ala Asn Arg Asn Thr Val Phe Phe Asp Ile Asn  
 165 170 175  
 20 Met Lys Gly Leu Asp Gly Ile Gln Gly Pro Val Tyr Val Gly Thr Gly  
 180 185 190  
 Cys Val Phe Arg Arg Gln Ala Leu Tyr Gly Tyr Glu Pro Pro Lys Gly  
 195 200 205  
 25 Pro Lys Arg Pro Lys Met Val Thr Cys Gly Cys Cys Pro Cys Phe Gly  
 210 215 220  
 Arg Arg Arg Lys Asp Lys Lys His Ser Lys Asp Gly Gly Asn Ala Asn  
 225 230 235 240  
 30 Gly Leu Ser Leu Glu Ala Ala Xaa Asp Asp Lys Glu Leu Leu Met Ser  
 245 250 255  
 His Met Asn Phe Glu Lys Lys Phe Gly Gln Ser Ala Ile Phe Val Thr  
 260 265 270  
 35 Ser Thr Leu Met Glu Gln Gly Gly Val Pro Pro Ser Ser Ser Pro Ala  
 275 280 285  
 Ala Leu Leu Lys Glu Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp  
 290 295 300  
 40 Lys Thr Glu Trp Gly Ser Glu Leu Gly Trp Ile Tyr Gly Ser Ile Thr  
 305 310 315 320  
 Glu Asp Ile Leu Thr Gly Phe Lys Met His Cys Arg Gly Trp Arg Ser  
 325 330 335  
 45 Ile Tyr Cys Met Pro Lys Leu Pro Ala Phe Lys Gly Ser Ala Pro Ile  
 340 345 350  
 Asn Leu Ser Asp Arg Leu Asn Gln Val Leu Arg Trp Ala Leu Gly Ser  
 355 360 365  
 50 Val Glu Ile Phe Phe Ser His His Cys Pro Ala Trp Tyr Gly Phe Lys  
 370 375 380  
 55 Gly Gly Lys Leu Lys Trp Leu Glu Arg Phe Ala Tyr Val Asn Thr Thr

385                                      390                                      395                                      400  
 Ile Tyr Pro Phe Thr Ser Leu Pro Leu Leu Ala Tyr Cys Thr Leu Pro  
 5                                      405                                      410                                      415  
 Ala Ile Cys Leu Leu Thr Asp Lys Phe Il Met Pro Pro Ile Ser Thr  
 420                                      425                                      430  
 Phe Ala Ser Leu Phe Phe Ile Ala Leu Phe Leu Ser Ile Phe Ala Thr  
 10                                      435                                      440                                      445  
 Gly Ile Leu Glu Leu Arg Trp Ser Gly Val Ser Ile Glu Glu Trp Trp  
 450                                      455                                      460  
 Arg Asn Glu Gln Phe Trp Val Ile Gly Gly Ile Ser Ala His Leu Phe  
 15                                      465                                      470                                      475                                      480  
 Ala Val Ile Gln Gly Leu Leu Lys Val Leu Ala Gly Ile Asp Thr Asn  
 485                                      490                                      495  
 Phe Thr Val Thr Ser Lys Ala Thr Asp Asp Glu Glu Phe Gly Glu Leu  
 20                                      500                                      505                                      510  
 Tyr Thr Phe Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Val Leu  
 515                                      520                                      525  
 Ile Ile Asn Leu Val Gly Val Val Ala Gly Ile Ser Asp Ala Ile Asn  
 25                                      530                                      535                                      540  
 Asn Gly Tyr Gln Ser Trp Gly Pro Leu Phe Gly Lys Leu Phe Phe Ser  
 545                                      550                                      555                                      560  
 30 Phe Trp Val Ile Val His Leu Tyr Pro Phe Leu Lys Gly Leu Met Gly  
 565                                      570                                      575  
 Arg Gln Asn Arg Thr Pro Thr Ile Val Val Ile Trp Ser Val Leu Leu  
 580                                      585                                      590  
 35 Ala Ser Ile Phe Ser Leu Leu Trp Val Arg Ile Asp Pro Phe Val Met  
 595                                      600                                      605  
 Lys Thr Lys Gly Pro Asp Thr Thr Met Cys Gly Ile Asn Cys  
 40                                      610                                      615                                      620

## (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Gln Xaa Xaa Xaa Xaa Xaa Xaa Arg Trp

1

5

## (2) INFORMATION FOR SEQ ID NO: 13:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic DNA"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GAGAGAGAGA GAGAGAGAGA ACTAGTCTOG AGTTTTTTTTT TTTTTTTTTT

50

## (2) INFORMATION FOR SEQ ID NO: 14:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic DNA"

## (ix) FEATURE:

- (A) NAME/KEY:
- (B) LOCATION:1..4
- (D) OTHER INFORMATION: single strand

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AATTCGGCAC GAG

13

## (2) INFORMATION FOR SEQ ID NO: 15:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic DNA"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GACTGAAGAT AAGCCAAAAG

20

## (2) INFORMATION FOR SEQ ID NO: 16:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGAATGATGA ATTTGCOGG

19

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGCAGGCAAC TTTGGCATGC

20

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AGCAACACGA GCAAGATGAG GAGGATGACT

30

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

COGGATCCTT CAACCTTCT TOGATTTC

28

(2) INFORMATION FOR SEQ ID NO: 20:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

COGGATOCAC GGCAATGCAT CTTGAAACC

29

## (2) INFORMATION FOR SEQ ID NO: 21:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GGTTAGCATA TTGTTTGTAG CATTGGG

27

## (2) INFORMATION FOR SEQ ID NO: 22:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ATCAATGAAA TATGTATAGT TCATAGC

27

## (2) INFORMATION FOR SEQ ID NO: 23:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CPTTGGTCT TTTGGTTTTG CCATGGC

27

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AGACTTTTGA CAAACAAGAT AAATCC

27

Claims

1. A DNA coding for any one of the following proteins (A) to (C):

(A) a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO: 2 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO: 2;

(B) a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO: 4 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO: 4; and

(C) a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO: 8 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO: 8, and an amino acid sequence shown in SEQ ID NO: 11 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO: 11.

2. A recombinant vector comprising all or a part of the DNA as defined in claim 1.

3. A transformed cell transformed with the DNA as defined in claim 1.

4. A method for controlling cellulose synthesis in a cell, comprising the steps of introducing the DNA as defined in claim 1 into the cell, and expressing RNA having a nucleotide sequence homologous to the DNA as defined in claim 1 or a nucleotide sequence complementary to the DNA as defined in claim 1.



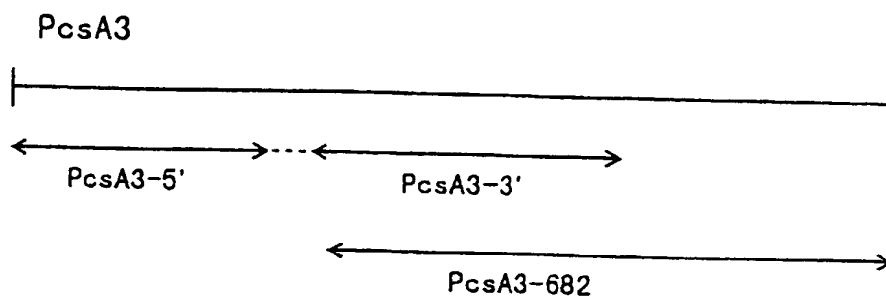


FIG. 1

SEQ ID NO: 14

```

5'  AATTCGGCACGAG  3'
3'          GCCGTGCTC  5' ---
    
```

FIG. 2

	10	20	30	40	50	60
PcsA3-682	CCGACATT	CGTAAGGAGCGT	CGAGCTAT	GAAGAGAGAAT	TGAAGAATT	CAAGGTTAGG
(SEQ ID NO: 5)	.....					
PcsA3-3'	CCGACATT	CGTAAGGAGCGT	CGAGCTAT	GAAGAGAGAAT	TGAAGAATT	CAAGGTTAGG
(SEQ ID NO: 9)	20	30	40	50	60	70
	70	80	90	100	110	120
PcsA3-682	ATAAATGCACTT	GTAGCCAAAGCCCAA	AAAGGTTCTCCAGA	AGGGTGGATCAT	GCAAGAT	
	.....					
PcsA3-3'	ATAAATGCACTT	GTAGCCAAAGCCCAA	AAAGGTTCTCCAGA	AGGGTGGATCAT	GCAAGAT	
	80	90	100	110	120	130
	130	140	150	160	170	180
PcsA3-682	GGGACACCAT	GGCCAGGAAACA	ATACTAAAGAT	CACCCCTGGTAT	GATTCAAGTAT	TTTCTC
	.....					
PcsA3-3'	GGGACACCAT	GGCCAGGAAACA	ATACTAAAGAT	CACCCCTGGTAT	GATTCAAGTAT	TTTCTC
	140	150	160	170	180	190
	190	200	210	220	230	240
PcsA3-682	GGTCAAAGT	GGAGGCCATGAT	ACCGAAGGAAAT	GAGCTTCCTCGT	CTCGTCTATG	TATCT
	.....					
PcsA3-3'	GGTCAAAGT	GGAGGCCATGAT	ACCGAAGGAAAT	GAGCTTCCTCGT	CTCGTCTATG	TATCT
	200	210	220	230	240	250
	250	260	270	280	290	300
PcsA3-682	CGAGAGAAAAGG	CCTGGTTTCTTGC	ATCACAAGAAAGCT	GGTGCCATGA	ACGCCCTTGT	
	.....					
PcsA3-3'	CGAGAGAAAAGG	CCTGGTTTCTTGC	ATCACAAGAAAGCT	GGTGCCATGA	ACGCCCTTGT	
	260	270	280	290	300	310
	310	320	330	340	350	360
PcsA3-682	CGGGTCTCGGGGG	TGCTCACAAATGCT	CCTTTTATGTTGA	ACTTGGATTGT	GACCAT	
	.....					
PcsA3-3'	CGGTCTCGGGGG	TGCTTACAAATGCT	CCTTTTATGTTGA	ACTTGGATTGT	GACCAT	
	320	330	340	350	360	370
	370	380	390	400	410	420
PcsA3-682	TTAAATAACAGCA	AGGCTGTAAGAGAGG	CTATGTGTTTCTT	GATGGACCC	CAAATTGGA	
	.....					
PcsA3-3'	TTAAATAACAGCA	AGGCTGTAAGAGAGG	CTATGTGTTTCTT	GATGGACCC	CAAATTGGA	
	380	390	400	410	420	430
	430	440	450	460	470	480
PcsA3-682	AGAAAGGTTTGCT	ATGTCCAATTCCT	CAACGTTTCGAT	GGTATTGATAG	ACATGATCGA	
	.....					
PcsA3-3'	AGAAAGGTTTGCT	ATGTCCAATTCCT	CAACGTTTCGAT	GGTATTGATAG	ACATGATCGA	
	440	450	460	470	480	490
	490	500	510	520	530	540
PcsA3-682	TATGCCAATCGGA	ACACAGTTTTCTTT	GATTTAACATGAA	AGGTCTAGAT	GGTATACAA	
	.....					
PcsA3-3'	TATGCCAATCGGA	ACACAGTTTTCTTT	GATTTAACATGAA	AGGTCTAGAT	GGTATACAA	
	500	510	520	530	540	550

FIG. 3

	550	560	570	580	590	600
PcsA3-682	GGCCCTGTATATGTCGGCACGGGGTGTGTTTTTCAGAAGGCAAGCTCTTTATGGTTATGAA					
(SEQ ID NO: 5)	.....	.....	.....	.....	.....	.....
PcsA3-3'	GGCCCTGTATATGTCGGCACGGGGTGTGTTTTTCAGAAGGCAAGCTCTTTATGGTTATGAA					
(SEQ ID NO: 9)	.....	.....	.....	.....	.....	.....
	560	570	580	590	600	610
	610	620	630	640	650	660
PcsA3-682	CCTCCAAAGGGACCTAAGCGCCCGAAAAATGGTAACCTGTGGTTGCTGCCCTTGTTTTGGGA					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	CCTCCAAAGGGACCTAAGCGCCCGAAAAATGGTAACCTGTGGTTGCTGCCCTTGTTTTGGGA					
	.....	.....	.....	.....	.....	.....
	620	630	640	650	660	670
	670	680	690	700	710	720
PcsA3-682	CGCCGCAGAAAGGACAAAAAGCACTCTAAGGATGGTGGAAATGCAAATGGTCTAAGCCTA					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	CGCCGCAGAAAGGACAAAAAGCACTCTAAGGATGGTGGAAATGCAAATGGTCTAAGCCTA					
	.....	.....	.....	.....	.....	.....
	680	690	700	710	720	730
	730	740	750	760	770	780
PcsA3-682	GAAGCAGCCAAAGATGACAAGGAGTTATTGATGTCCACATGAACTTTGAAAAGAAATTT					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	GAAGCAGCCGAAGATGACAAGGAGTTATTGATGTCCACATGAACTTTGAAAAGAAATTT					
	.....	.....	.....	.....	.....	.....
	740	750	760	770	780	790
	790	800	810	820	830	840
PcsA3-682	GGACAATCAGCCATTTTGTAACTTCAACACTGATGGAACAAGGTGGTGTCCCTCCTTCT					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	GGACAATCAGCCATTTTGTAACTTCAACACTGATGGAACAAGGTGGTGTCCCTCCTTCT					
	.....	.....	.....	.....	.....	.....
	800	810	820	830	840	850
	850	860	870	880	890	900
PcsA3-682	TCAAGCCCCGCAGCTTTGCTCAAAGAAGCCATTATGTAATTAGTTGTGGTTATGAAGAC					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	TCAAGCCCTGCAGCTTTGCTCAAAGAAGCCATTATGTAATTAGTTGTGGTTATGAAGAC					
	.....	.....	.....	.....	.....	.....
	860	870	880	890	900	910
	910	920	930	940	950	960
PcsA3-682	AAAACAGAATGGGGAAGCGAGCTTGCTGGATTACGGGTCGATTACAGAAGATATCTTA					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	AAAACCGAATGGGGAAGCGAGCTTGCTGGATTACGGGTCGATTACAGAAGATATCTTA					
	.....	.....	.....	.....	.....	.....
	920	930	940	950	960	970
	970	980				
PcsA3-682	ACAGGATTCAAGATGCATTGCCGTGGAT					
	.....	.....	.....	.....	.....	.....
PcsA3-3'	ACAGGTTTCAAGATGCATTGCCGTGGAT					
	.....	.....	.....	.....	.....	.....
	980	990	1000			

FIG. 4



(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 875 575 A3**

(12)

**EUROPEAN PATENT APPLICATION**

(88) Date of publication A3:  
**30.06.1999 Bulletin 1999/26**

(51) Int Cl.<sup>6</sup>: **C12N 15/54, C12N 9/10**

(43) Date of publication A2:  
**04.11.1998 Bulletin 1998/45**

(21) Application number: **98302489.4**

(22) Date of filing: **31.03.1998**

(84) Designated Contracting States:  
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**  
Designated Extension States:  
**AL LT LV MK RO SI**

- Aotsuka, Satoshi c/o Nisshinbo Ind., Inc.,  
Adachi-ku, Tokyo (JP)
- Ihara, Yuri  
Takaishi, Osaka (JP)
- Hayashi, Takahisa  
Uji Kyoto 611 (JP)

(30) Priority: **01.04.1997 JP 8313397**

(71) Applicant: **NISSHINBO INDUSTRIES, INC.**  
**Chuo-ku, Tokyo 103 (JP)**

(74) Representative: **Bannerman, David Gardner et al**  
**Withers & Rogers,**  
**Goldings House,**  
**2 Hays Lane**  
**London SE1 2HW (GB)**

(72) Inventors:  
• **Hasegawa, Osamu, c/o Nisshinbo Ind., Inc.**  
**Adachi-ku, Tokyo (JP)**

**(54) Cellulose synthase gene**

(57) mRNA was extracted at the stage for cotton plant fibrous cells to accumulate cellulose, and cDNA's complementary thereto were synthesized to construct a cDNA library. Clones of a number of 750 were arbitrarily selected from the library, and they were randomly subjected from to sequencing. Those having homology to

an amino acid sequence deduced from a gene of cellulose 4- $\beta$ -glucosyltransferase (bcsA) of cellulose synthase operon of acetic acid bacterium were selected from obtained nucleotide sequences of the respective clones. Thus, DNA coding for cellulose synthase was obtained.

**EP 0 875 575 A3**



European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number  
EP 98 30 2489

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	WO 98 00549 A (THE AUSTRALIAN NATIONAL UNIVERSITY; COMMONWEALTH SCIENTIFIC...) 8 January 1998 * page 1, line 3 - line 11 * * page 2, line 21 - page 7, line 28 * * example 8 * 'Sequence Listing: SEQ ID NO.9 and 10' ---	1-4	C12N15/54 C12N9/10
X,D	PEAR, J.R. ETAL.: "Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase" PROC.NATL.ACAD.SCI.USA, vol. 93, October 1996, pages 12637-12642, XP002061424 * the whole document * ---	1-4	
Y	WO 91 13988 A (THE BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM) 19 September 1991 * page 1, line 18 - line 26 * * page 5, line 15 - page 8, line 13 * * figure 1; examples I,II,IV,V * ---	1-4	TECHNICAL FIELDS SEARCHED (Int.Cl.6) C12N
Y	LI, L. ET AL.: "B-Glucan synthesis in the cotton fiber" PLANT PHYSIOLOGY, vol. 101, no. 4, 1993, pages 1149-1156, XP002087180 * page 1149 * * page 1154 - page 1155 * 'Abstract' and 'Discussion' ---	1-4	
E	WO 98 18949 A (CALGENE, INC.) 7 May 1998 * page 7, line 14 - page 9, line 25 * * figures 3,6,8; examples 1-7 * -----	1-4	
The present search report has been drawn up for all claims			
Place of search MUNICH		Date of completion of the search 8 December 1998	Examiner Donath, C
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X: particularly relevant if taken alone Y: particularly relevant if combined with another document of the same category A: technological background O: non-written disclosure P: intermediate document</p> <p>T: theory or principle underlying the invention E: earlier patent document, but published on, or after the filing date D: document cited in the application L: document cited for other reasons &amp;: member of the same patent family, corresponding document</p>			

Environ Biol Fish (2003) 62:153-154



European Patent  
Office

Application Number  
EP 98 30 2489

**CLAIMS INCURRING FEES**

The present European patent application comprised at the time of filing more than ten claims.

- ☐ Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid, namely claim(s):
- ☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

**LACK OF UNITY OF INVENTION**

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

see sheet B

- ☐ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.
- ☐ As all searchable claims could be searched without effort justifying an additional fee, the Search Division did not invite payment of any additional fee.
- ☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:
- ☒ None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:
- 1-4 (partially)



European Patent  
Office

LACK OF UNITY OF INVENTION  
SHEET B

Application Number

EP 98 30 2489

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. Claims: 1-4 (partially)

Claims 1- 4 (partially) refer to a DNA coding for a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO:2 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO:2, a recombinant vector comprising all or part of said DNA, a cell being transformed with said DNA, and a method for controlling cellulose synthesis in a cell by the use of said DNA.

2. Claims: 1-4 (partially)

Claims 1- 4 (partially) refer to a DNA coding for a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO:4 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO:4, a recombinant vector comprising all or part of said DNA, a cell being transformed with said DNA, and a method for controlling cellulose synthesis in a cell by the use of said DNA.

3. Claims: 1-4 (partially)

Claims 1- 4 (partially) refer to a DNA coding for a protein having a cellulose synthase activity and comprising an amino acid sequence shown in SEQ ID NO:8 and in SEQ ID NO:11 or an amino acid sequence involving deletion, substitution, insertion, or addition of one or several amino acids relevant to SEQ ID NO:8 and/or SEQ ID NO:11, a recombinant vector comprising all or part of said DNA, a cell being transformed with said DNA, and a method for controlling cellulose synthesis in a cell by the use of said DNA.